(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 September 2003 (25.09.2003)

PCT

(10) International Publication Number
WO 03/078577 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/07430

(22) International Filing Date: 12 March 2003 (12.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/363,684 12 March 2002 (12.03.2002) US

(71) Applicant: U.S. SMOKELESS TOBACCO COMPANY [US/US]; 100 West Putnam Avenue, Greenwich, CT 06830 (US).

(72) Inventor: XU, Dongmei; Lexington, KY (US).

(74) Agents: SAMPLES, Kenneth, H. et al.; Fitch, Even, Tabin & Flannery, 120 South LaSalle Street, Suite 1600, Chicago, IL 60603 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/078577 A2

(54) Title: CLONING OF CYTOCHROME P450 GENES FROM NICOTIANA

(57) Abstract: The present invention relates to P450 enzymes and nucleic acid sequences encoding P450 enzymes in Nicotiana, and methods of using those enzymes and nucleic acid sequences to alter plant phenotypes.

CLONING OF CYTOCHROME P450 GENES FROM NICOTIANA

The present invention relates to nucleic acid sequences encoding cytochrome P450 enzymes (hereinafter referred to as P450 and P450 enzymes) in *Nicotiana* plants and methods for using those nucleic acid sequences to alter plant phenotypes.

BACKGROUND

Cytochrome P450s catalyze enzymatic reactions for a diverse range of chemically dissimilar substrates that include the oxidative, peroxidative and reductive metabolism of endogenous and xenobiotic substrates. In plants, P450s participate in biochemical pathways that include the synthesis of plant products such as phenylpropanoids, alkaloids, terpenoids, lipids, cyanogenic glycosides, and glucosinolates (Chappel, Annu. Rev. Plant Physiol. Plant Mol. Biol. 198, 49:311-343). Cytochrome P450s, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450- containing monooxygenase systems. Specific reactions catalyzed include demethylation, hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O- dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups.

The diverse role of *Nicotiana* plant P450 enzymes has been implicated in effecting a variety of plant metabolites such as phenylpropanoids, alkaloids, terpenoids, lipids, cyanogenic glycosides, glucosinolates and a host of other chemical entities. During recent years, it is becoming apparent that some P450 enzymes can impact the composition of plant metabolites in plants. For example, it has been long desired to improve the flavor and aroma of certain plants by altering its profile of selected fatty acids through breeding; however

very little is known about mechanisms involved in controlling the levels of these leaf constituents. The down regulation of P450 enzymes associated with the modification of fatty acids may facilitate accumulation of desired fatty acids that provide more preferred *leaf* phenotypic qualities. The function of P450 enzymes and their broadening roles in plant constituents is still being discovered. For instance, a special class of P450 enzymes was found to catalyze the breakdown of fatty acid into volatile C6- and C9-aldehydes and -alcohols that are major contributors of "fresh green" odor of fruits and vegetables. The level of other novel targeted P450s may be altered to enhance the qualities of leaf constituents by modifying lipid composition and related break down metabolites in Nicotiana leaf. Several of these constituents in leaf are affected by senescence that stimulates the maturation of leaf quality properties. Still other reports have shown that P450s enzymes are play a functional role in altering fatty acids that are involved in plant-pathogen interactions and disease resistance.

In other instances, P450 enzymes have been suggested to be involved in alkaloid biosynthesis. Nornicotine is a minor alkaloid found in *Nicotiana tabaccum*. It is has been postulated that it is produced by the P450 mediated demethylation of nicotine followed by acylation and nitrosation at the N position thereby producing a series of N-acylnonicotines and N-nitrosonornicotines. N-demethylation, catalyzed by a putative P450 demethylase, is thought to be a primary source of nornicotine biosyntheses in *Nicotiana*. While the enzyme is believed to be microsomal, thus far a nicotine demethylase enzyme has not been successfully purified, nor have the genes involved been isolated.

Furthermore, it is hypothesized but not proven that the activity of P450 enzymes is genetically controlled and also strongly influenced by environment factors. For example, the

demethylation of nicotine in *Nicotiana* is thought to increase substantially when the plants reach a mature stage. Furthermore, it is thought that the demethylase gene contains a transposable element that can inhibit translation of RNA when present.

The large multiplicity of P450 enzyme forms, their differing structure and function have made their research on *Nicotiana* P450 enzymes very difficult before the enclosed invention. In addition, cloning of P450 enzymes has been hampered at least in part because these membrane-localized proteins are typically present in low abundance and often unstable to purification. Hence, a need exists for the identification of P450 enzymes in plants and the nucleic acid sequences associated with those P450 enzymes. In particular, only a few cytochromé Nicotiana P450 proteins have been reported in tobacco. The inventions described herein entail the discovery of a substantial number of cytochrome P450 fragments that correspond to several groups of P450 species based on their sequence identity.

SUMMARY

The present invention is directed to plant P450 enzymes. The present invention is further directed to plant P450 enzymes from *Nicotiana*. The present invention is also directed to P450 enzymes in plants whose expression is induced by ethylene and/or plant senescence. The present invention is yet further directed to nucleic acid sequences in plants having enzymatic activities, for example, oxygenase, demethylase and the like, or other and the use of those sequences to reduce or silence the expression of these enzymes. The invention also relates to P450 enzymes found in plants containing higher nornicotine levels than plants exhibiting lower nornicotine levels.

In one aspect, the invention is directed to nucleic acid sequences as set forth in SEQ. ID. Nos. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145 and 147.

In a second related aspect, those fragments containing greater than 75% identity in nucleic acid sequence were placed into groups dependent upon their identity in a region corresponding to the first nucleic acid following the cytochrome P450 motif GXRXCX(G/A) to the stop codon. The representative nucleic acid groups and respective species are shown in Table I.

In a third aspect, the invention is directed to amino acid sequences as set forth in SEQ. ID. Nos. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146 and 148.

In a fourth related aspect, those fragments containing greater than 71% identity in amino acid sequence were placed into groups dependent upon their identity to each other in a region corresponding to the first amino acid following the cytochrome P450 motif GXRXCX(G/A) to the stop codon. The representative amino acid groups and respective species are shown in Table II.

In a fifth aspect of the invention is the use of nucleic acids sequences as set forth in SEQ. ID. Nos. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41,

43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145 and 147.

In a sixth related aspect, the reduction or elimination of P450 enzymes in Nicotiana plants may be accomplished transiently using RNA viral systems. Resulting transformed or infected plants are assessed for phenotypic changes including, but not limited to, analysis of endogenous P450 RNA transcripts, P450 expressed peptides, and concentrations of plant metabolites using techniques commonly available to one having ordinary skill in the art.

In a seventh important aspect, the present invention is also directed to generation of trangenic Nicotiana lines that have altered P450 enzyme activity levels. In accordance with the invention, these transgenic lines include nucleic acid sequences that are effective for reducing or silencing the expression of certain enzyme thus resulting in phenotypic effects within Nicotiana. Such nucleic acid sequences include SEQ. ID. Nos. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145 and 147.

In a very important eight aspect of the invention, plant cultivars including nucleic acids of the present invention in a down regulation capacity will have altered metabolite profiles relative to control plants.

In a ninth aspect, the present invention is directed to the screening of plants, more preferably Nicotiana, that

contain genes that have substantial nucleic acid identity to the taught nucleic acid sequence. The use of the invention would be advantageous to identify and select plants that contain a nucleic acid sequence with exact or substantial identity where such plants are part of a breeding program for traditional or transgenic varieties, a mutagenesis program, or naturally occurring diverse plant populations. The screening of plants for substantial nucleic acid identity may be accomplished by evaluating plant nucleic acid materials using a nucleic acid probe in conjunction with nucleic acid detection protocols including, but not limited to, nucleic acid hybridization and PCR analysis. The nucleic acid probe may consist of the taught nucleic acid sequence or fragment thereof corresponding to SEQ ID 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145 and 147.

In a tenth aspect, the present invention is directed to the identification of plant genes, more preferably Nicotiana, that share substantial amino acid identity corresponding to the taught nucleic acid sequence. The identification of plant genes including both cDNA and genomic clans of those cDNAs and genomic clones, preferably from Nicotiana may be accomplished by screening plant cDNA libraries using a nucleic acid probe in conjunction with nucleic acid detection protocols including, but not limited to, nucleic acid hybridization and PCR analysis. The nucleic acid probe may be comprised of nucleic acid sequence or fragment thereof corresponding to SEQ ID 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99,

101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145 and 147

In an alterative eleventh aspect, cDNA expression libraries that express peptides may be screened using antibodies directed to part or all of the taught amino acid sequence. Such amino acid sequences include SEQ ID 2, 4, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146 and 148.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows nucleic acid SEQ. ID. No.:1 and amino acid SEQ. ID. No.:2.

Figure 2 shows nucleic acid SEQ. ID. No.:3 and amino acid SEQ. ID. No.:4.

Figure 3 shows nucleic acid SEQ. ID. No.:5 and amino acid SEQ. ID. No.:6.

Figure 4 shows nucleic acid SEQ. ID. No.:7 and amino acid SEQ. ID. No.:8.

Figure 5 shows nucleic acid SEQ. ID. No.:9 and amino acid SEQ. ID. No.:10.

Figure 6 shows nucleic acid SEQ. ID. No.:11 and amino acid SEQ. ID. No.:12.

Figure 7 shows nucleic acid SEQ. ID. No.:13 and amino acid SEQ. ID. No.:14.

Figure 8 shows nucleic acid SEQ. ID. No.:15 and amino acid SEQ. ID. No.:16.

Figure 9 shows nucleic acid SEQ. ID. No.:17 and amino acid SEQ. ID. No.:18.

Figure 10 shows nucleic acid SEQ. ID. No.:19 and amino acid SEQ. ID. No.:20.

Figure 11 shows nucleic acid SEQ. ID. No.:21 and amino acid SEQ. ID. No.:22.

Figure 12 shows nucleic acid SEQ. ID. No.:23 and amino acid SEQ. ID. No.:24.

Figure 13 shows nucleic acid SEQ. ID. No.:25 and amino acid SEQ. ID. No.:26.

Figure 14 shows nucleic acid SEQ. ID. No.:27 and amino acid SEQ. ID. No.:28.

Figure 15 shows nucleic acid SEQ. ID. No.:29 and amino acid SEQ. ID. No.:30.

Figure 16 shows nucleic acid SEQ. ID. No.:31 and amino acid SEQ. ID. No.:32.

Figure 17 shows nucleic acid SEQ. ID. No.:33 and amino acid SEQ. ID. No.:34.

Figure 18 shows nucleic acid SEQ. ID. No.:35 and amino acid SEQ. ID. No.:36.

Figure 19 shows nucleic acid SEQ. ID. No.:37 and amino acid SEQ. ID. No.:38.

Figure 20 shows nucleic acid SEQ. ID. No.:39 and amino acid SEQ. ID. No.:40.

Figure 21 shows nucleic acid SEQ. ID. No.:41 and amino acid SEQ. ID. No.:42.

Figure 22 shows nucleic acid SEQ. ID. No.:43 and amino acid SEQ. ID. No.:44.

Figure 23 shows nucleic acid SEQ. ID. No.:45 and amino acid SEQ. ID. No.:46.

Figure 24 shows nucleic acid SEQ. ID. No.:47 and amino acid SEQ. ID. No.:48.

Figure 25 shows nucleic acid SEQ. ID. No.:49 and amino acid SEQ. ID. No.:50.

Figure 26 shows nucleic acid SEQ. ID. No.:51 and amino acid SEQ. ID. No.:52.

Figure 27 shows nucleic acid SEQ. ID. No.:53 and amino acid SEQ. ID. No.:54.

Figure 28 shows nucleic acid SEQ. ID. No.:55 and amino acid SEQ. ID. No.:56.

Figure 29 shows nucleic acid SEQ. ID. No.:57 and amino acid SEQ. ID. No.:58.

Figure 30 shows nucleic acid SEQ. ID. No.:59 and amino acid SEQ. ID. No.:60.

Figure 31 shows nucleic acid SEQ. ID. No.:61 and amino acid SEQ. ID. No.:62.

Figure 32 shows nucleic acid SEQ. ID. No.:63 and amino acid SEQ. ID. No.:64.

Figure 33 shows nucleic acid SEQ. ID. No.:65 and amino acid SEQ. ID. No.:66.

Figure 34 shows nucleic acid SEQ. ID. No.:67 and amino acid SEQ. ID. No.:68.

Figure 35 shows nucleic acid SEQ. ID. No.:69 and amino acid SEQ. ID. No.:70.

Figure 36 shows nucleic acid SEQ. ID. No.:71 and amino acid SEQ. ID. No.:72.

Figure 37 shows nucleic acid SEQ. ID. No.:73 and amino acid SEQ. ID. No.:74.

Figure 38 shows nucleic acid SEQ. ID. No.:75 and amino acid SEQ. ID. No.:76.

Figure 39 shows nucleic acid SEQ. ID. No.:77 and amino acid SEQ. ID. No.:78.

Figure 40 shows nucleic acid SEQ. ID. No.:79 and amino acid SEQ. ID. No.:80.

Figure 41 shows nucleic acid SEQ. ID. No.:81 and amino acid SEQ. ID. No.:82.

Figure 42 shows nucleic acid SEQ. ID. No.:83 and amino acid SEQ. ID. No.:84.

Figure 43 shows nucleic acid SEQ. ID. No.:85 and amino acid SEQ. ID. No.:86.

Figure 44 shows nucleic acid SEQ. ID. No.:87 and amino acid SEQ. ID. No.:88.

Figure 45 shows nucleic acid SEQ. ID. No.:89 and amino acid SEQ. ID. No.:90.

Figure 46 shows nucleic acid SEQ. ID. No.:91 and amino acid SEQ. ID. No.:92.

Figure 47 shows nucleic acid SEQ. ID. No.:93 and amino acid SEQ. ID. No.:94.

Figure 48 shows nucleic acid SEQ. ID. No.:95 and amino acid SEQ. ID. No.:96.

Figure 49 shows nucleic acid SEQ. ID. No.:97 and amino acid SEQ. ID. No.:98.

Figure 50 shows nucleic acid SEQ. ID. No.:99 and amino acid SEQ. ID. No.:100.

Figure 51 shows nucleic acid SEQ. ID. No.:101 and amino acid SEQ. ID. No.:102.

Figure 52 shows nucleic acid SEQ. ID. No.:103 and amino acid SEQ. ID. No.:104.

Figure 53 shows nucleic acid SEQ. ID. No.:105 and amino acid SEQ. ID. No.:106.

Figure 54 shows nucleic acid SEQ. ID. No.:107 and amino acid SEQ. ID. No.:108.

Figure 55 shows nucleic acid SEQ. ID. No.:109 and amino acid SEQ. ID. No.:110.

Figure 56 shows nucleic acid SEQ. ID. No.:111 and amino acid SEQ. ID. No.:112.

Figure 57 shows nucleic acid SEQ. ID. No.:113 and amino acid SEQ. ID. No.:114.

Figure 58 shows nucleic acid SEQ. ID. No.:115 and amino acid SEQ. ID. No.:116.

Figure 59 shows nucleic acid SEQ. ID. No.:117 and amino acid SEQ. ID. No.:118.

Figure 60 shows nucleic acid SEQ. ID. No.:119 and amino acid SEQ. ID. No.:120.

Figure 61 shows nucleic acid SEQ. ID. No.:121 and amino acid SEQ. ID. No.:122.

Figure 62 shows nucleic acid SEQ. ID. No.:123 and amino acid SEQ. ID. No.:124.

Figure 63 shows nucleic acid SEQ. ID. No.:125 and amino acid SEQ. ID. No.:126.

Figure 64 shows nucleic acid SEQ. ID. No.:127 and amino acid SEQ. ID. No.:128.

Figure 65 shows nucleic acid SEQ. ID. No.:129 and amino acid SEQ. ID. No.:130.

Figure 66 shows nucleic acid SEQ. ID. No.:131 and amino acid SEQ. ID. No.:132.

Figure 67 shows nucleic acid SEQ. ID. No.:133 and amino acid SEQ. ID. No.:134.

Figure 68 shows nucleic acid SEQ. ID. No.:135 and amino acid SEQ. ID. No.:136.

Figure 69 shows nucleic acid SEQ. ID. No.:137 and amino acid SEQ. ID. No.:138.

Figure 70 shows nucleic acid SEQ. ID. No.:139 and amino acid SEQ. ID. No.:140.

Figure 71 shows nucleic acid SEQ. ID. No.:141 and amino acid SEQ. ID. No.:142.

Figure 72 shows nucleic acid SEQ. ID. No.:143 and amino acid SEQ. ID. No.:144.

Figure 73 shows nucleic acid SEQ. ID. No.:145 and amino acid SEQ. ID. No.:146.

Figure 74 shows nucleic acid SEQ. ID. No.:147 and amino acid SEQ. ID. No.:148.

Figure 75 shows a procedure used for cloning of cytochrome P450 cDNA fragments by PCR. SEQ. ID. Nos. 149-156 are shown.

Figure 76 illustrates amino acid identity of group members.

DETAILED DESCRIPTION

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood

by one of ordinary skill in the art to which this invention belongs. Singleton et al. (1994) Dictionary of Microbiology and Molecular Biology, second edition, John Wiley and Sons (New York) provides one of skill with a general dictionary of many of the terms used in this invention. All patents and publications referred to herein are incorporated by reference herein. For purposes of the present invention, the following terms are defined below.

"Enzymatic activity" is meant to include demethylation, hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O- dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, or sense or anti-sense, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof. The terms "operably linked", "in operable combination", and "in operable order" refer to functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence affects transcription and/or translation of the nucleic acid corresponding to the second sequence.

The term "recombinant" when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, expresses said nucleic acid or expresses a peptide, heterologous peptide, or protein encoded by a heterologous nucleic acid. Recombinant cells can express genes or gene fragments in either the sense or antisense form that are not found within the native (non- recombinant) form of the cell.

Recombinant cells can also express genes that are found in the native form of the cell, but wherein the genes are modified and re- introduced into the cell by artificial means.

A "structural gene" is that portion of a gene comprising a DNA segment encoding a protein, polypeptide or a portion thereof, and excluding the 5' sequence which drives the initiation of transcription. The structural gene may alternatively encode a nontranslatable product. The structural gene may be one which is normally found in the cell or one which is not normally found in the cell or cellular location wherein it is introduced, in which case it is termed a "heterologous gene". A heterologous gene may be derived in whole or in part from any source known to the art, including a bacterial genome or episome, eukaryotic, nuclear or plasmid DNA, cDNA, viral DNA or chemically synthesized DNA. A structural gene may contain one or more modifications that could effect biological activity or its characteristics, the biological activity or the chemical structure of the expression product, the rate of expression or the manner of expression control. Such modifications include, but are not limited to, mutations, insertions, deletions and substitutions of one or more nucleotides. The structural gene may constitute an uninterrupted coding sequence or it may include one or more introns, bounded by the appropriate splice junctions. The structural gene may be translatable or non-translatable, including in an anti-sense orientation. The structural gene may be a composite of segments derived from a plurality of sources and from a plurality of gene sequences (naturally occurring or synthetic, where synthetic refers to DNA that is chemically synthesized).

"Derived from" is used to mean taken, obtained, received, traced, replicated or descended from a source (chemical and/or biological). A derivative may be produced by chemical or

biological manipulation (including, but not limited to, substitution, addition, insertion, deletion, extraction, isolation, mutation and replication) of the original source.

"Chemically synthesized", as related to a sequence of DNA, means that portions of the component nucleotides were assembled in vitro. Manual chemical synthesis of DNA may be accomplished using well established procedures (Caruthers, Methodology of DNA and RNA Sequencing, (1983), Weissman (ed.), Praeger Publishers, New York, Chapter 1); automated chemical synthesis can be performed using one of a number of commercially available machines.

Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson and Lipman Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) is available from several sources, including the National Center for Biological Information (NCBI, Bethesda, Md.) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. It can be accessed at htp://www.ncbi.nlm.nih.gov/BLAST/. A description of how to determine sequence identity using this program is available at http://www.ncbi.nlm.nih.gov/BLAST/blast help.html.

The terms "substantial amino acid identity" or "substantial amino acid sequence identity" as applied to amino acid sequences and as used herein denote a characteristic of a polypeptide, wherein the peptide comprises a sequence that has at least 70 percent sequence identity, preferably 80 percent amino acid sequence identity, more preferably 90 percent amino acid sequence identity, and most preferably at least 99 to 100 percent sequence identity as compared to a reference group over region corresponding to the first amino acid following the cytochrome P450 motif GXRXCX(G/A) to the stop codon of the translated peptide.

The terms "substantial nucleic acid identity" or "substantial nucleic acid sequence identity" as applied to nucleic acid sequences and as used herein denote a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 75 percent sequence identity, preferably 81 percent amino acid sequence identity, more preferably at least 91 to 99 percent sequence identity, and most preferably at least 99 to 100 percent sequence identity as compared to a reference group over region corresponding to the first nucleic acid following the cytochrome P450 motif GXRXCX(G/A) to the stop codon of the translated peptide.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C to about 20°C, usually about 10°C to about 15°C, lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a matched probe.

Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60°C. For instance, in a standard Southern hybridization procedure, stringent conditions will include an initial wash in 6xSSC at 42 °C followed by one or more additional washes in 0.2xSSC at a temperature of at least about 55°C, typically about 60°C and often about 65°C.

Nucleotide sequences are also substantially identical for purposes of this invention when the polypeptides and/or proteins which they encode are substantially identical. Thus, where one nucleic acid sequence encodes essentially the same polypeptide as a second nucleic acid sequence, the two nucleic acid sequences are substantially identical, even if they would not hybridize under stringent conditions due to degeneracy permitted by the genetic code (see, Darnell et al. (1990) Molecular Cell Biology, Second Edition Scientific American Books W. H. Freeman and Company New York for an explanation of codon degeneracy and the genetic code). Protein purity or homogeneity can be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualization upon staining. For certain purposes high resolution may be needed and HPLC or a similar means for purification may be utilized.

As used herein, the term "vector" is used in reference to nucleic acid molecules that transfer DNA segment(s) into a cell. A vector may act to replicate DNA and may reproduce independently in a host cell. The term "vehicle" is sometimes used interchangeably with "vector." The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes usually

include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eucaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

For the purpose of regenerating complete genetically engineered plants with roots, a nucleic acid may be inserted into plant cells, for example, by any technique such as in vivo inoculation or by any of the known in vitro tissue culture techniques to produce transformed plant cells that can be regenerated into complete plants. Thus, for example, the insertion into plant cells may be by in vitro inoculation by pathogenic or non-pathogenic *A. tumefaciens*. Other such tissue culture techniques may also be employed.

"Plant tissue" includes differentiated and undifferentiated tissues of plants, including, but not limited to, roots, shoots, leaves, pollen, seeds, tumor tissue and various forms of cells in culture, such as single cells, protoplasts, embryos and callus tissue. The plant tissue may be *in planta* or in organ, tissue or cell culture.

"Plant cell" as used herein includes plant cells *in planta* and plant cells and protoplasts in culture. "cDNA" or "complementary DNA" generally refers to a single stranded DNA molecule with a nucleotide sequence that is complementary to an RNA molecule. cDNA is formed by the action of the enzyme reverse transcriptase on an RNA template.

STRATEGIES FOR OBTAINING NUCLEIC ACID SEQUENCES

In accordance with the present invention, RNA was extracted from Nicotiana tissue of converter and non-converter Nicotiana lines. The extracted RNA was then used to create

cDNA. Nucleic acid sequences of the present invention were then generated using two strategies.

In the first strategy, the poly A enriched RNA was extracted from plant tissue and cDNA was made by reverse transcription PCR. The single strand cDNA was then used to create P450 specific PCR populations using degenerate primers plus a oligo d(T) reverse primer. The primer design was based on the highly conserved motifs of P450. Examples of specific degenerate primers are set forth in Figure 1. Sequence fragments from plasmids containing appropriate size inserts were further analyzed. These size inserts typically ranged from about 300 to about 800 nucleotides depending on which primers were used.

In a second strategy, a cDNA library was initially constructed. The cDNA in the plasmids was used to create p450 specific PCR populations using degenerate primers plus T7 primer on plasmid as reverse primer. As in the first strategy, sequence fragments from plasmids containing appropriate size inserts were further analyzed.

Nicotiana plant lines known to produce high levels of nornicotine (converter) and plant lines having undetectable levels of nornicotine may be used as starting materials.

Leaves can then be removed from plants and treated with ethylene to activate P450 enzymatic activities defined herein. Total RNA is extracted using techniques known in the art. cDNA fragments can then be generated using PCR (RT-PCR) with the oligo d(T) primer as described in Figure 1. The cDNA library can then be constructed more fully described in examples herein.

The conserved region of P450 type enzymes can be used as a template for degenerate primers (Figure 75). Using degenerate primers, P450 specific bands can be amplified by PCR. Bands indicative for P450 like enzymes can be identified by DNA sequencing. PCR fragments can be characterized using BLAST search, alignment or other tools to identify appropriate candidates.

Sequence information from identified fragments can be used to develop PCR primers. These primers are used to conduct quantitative RT-PCR from the RNA's of converter and non-converter ethylene treated plant tissue. Only appropriate sized DNA bands (300-800 bp) from converter lines or bands with higher density denoting higher expression in converter lines were used for further characterization. Large scale Southern reverse analysis were conducted to examine the differential expression for all clones obtained. In this aspect of the invention, these large scale reverse Southern assays can be conducted using labeled total cDNA's from different tissues as a probe to hybridize with cloned DNA fragments in order to screen all cloned inserts.

Nonradioactive Northern blotting assay was also used to characterize clones P450 fragments.

Nucleic acid sequences identified as described above can be examined by using virus induced gene silencing technology (VIGS, Baulcombe, Current Opinions in Plant Biology, 1999, 2:109-113).

In another aspect of the invention, interfering RNA technology (RNAi) is used to further characterize cytochrome P450 enzymatic activities in Nicotiana plants of the present invention. The following references which describe this technology are incorporated by reference herein, Smith et al.,

Nature, 2000, 407:319-320; Fire et al., Nature, 1998, 391:306-311; Waterhouse et al., PNAS, 1998, 95:13959-13964; Stalberg et al., Plant Molecular Biology, 1993, 23:671- 683; Baulcombe, Current Opinions in Plant Biology, 1999, 2:109-113; and Brigneti et al., EMBO Journal, 1998, 17(22):6739-6746. Plants may be transformed using RNAi techniques, antisense techniques, or a variety of other methods described.

Several techniques exist for introducing foreign genetic material into plant cells, and for obtaining plants that stably maintain and express the introduced gene. Such techniques include acceleration of genetic material coated onto microparticles directly into cells (US Patents 4,945,050 to Cornell and 5,141,131 to DowElanco). Plants may be transformed using Agrobacterium technology, see US Patent 5,177,010 to University of Toledo, 5,104,310 to Texas A&M, European Patent Application 0131624B1, European Patent Applications 120516, 159418B1, European Patent Applications 120516, 159418B1 and 176,112 to Schilperoot, US Patents 5,149,645, 5,469,976, 5,464,763 and 4,940,838 and 4,693,976 to Schilperoot, European Patent Applications 116718, 290799, 320500 all to MaxPlanck, European Patent Applications 604662 and 627752 to Japan Nicotiana, European Patent Applications 0267159, and 0292435 and US Patent 5,231,019 all to Ciba Geigy, US Patents 5,463,174 and 4,762,785 both to Calgene, and US Patents 5,004,863 and 5,159,135 both to Agracetus. Other transformation technology includes whiskers technology, see U.S. Patents 5,302,523 and 5,464,765 both to Zeneca. Electroporation technology has also been used to transform plants, see WO 87/06614 to Boyce Thompson Institute, 5,472,869 and 5,384,253 both to Dekalb, WO9209696 and WO9321335 both to PGS. All of these transformation patents and publications are incorporated by reference. In addition to numerous technologies for transforming plants, the type of tissue which is contacted with the foreign genes may vary as well. Such tissue would include

but would not be limited to embryogenic tissue, callus tissue type I and II, hypocotyl, meristem, and the like. Almost all plant tissues may be transformed during dedifferentiation using appropriate techniques within the skill of an artisan.

Foreign genetic material introduced into a plant may include a selectable marker. The preference for a particular marker is at the discretion of the artisan, but any of the following selectable markers may be used along with any other gene not listed herein which could function as a selectable marker. Such selectable markers include but are not limited to aminoglycoside phosphotransferase gene of transposon Tn5 (Aph II) which encodes resistance to the antibiotics kanamycin, neomycin and G418, as well as those genes which code for resistance or tolerance to glyphosate; hygromycin; methotrexate; phosphinothricin (bar); imidazolinones, sulfonylureas and triazolopyrimidine herbicides, such as chlorosulfuron; bromoxynil, dalapon and the like.

In addition to a selectable marker, it may be desirous to use a reporter gene. In some instances a reporter gene may be used without a selectable marker. Reporter genes are genes which are typically not present or expressed in the recipient organism or tissue. The reporter gene typically encodes for a protein which provide for some phenotypic change or enzymatic property. Examples of such genes are provided in K. Weising et al. Ann. Rev. Genetics, 22, 421 (1988), which is incorporated herein by reference. Preferred reporter genes include without limitation glucuronidase (GUS) gene and GFP genes.

Once introduced into the plant tissue, the expression of the structural gene may be assayed by any means known to the art, and expression may be measured as mRNA transcribed, protein synthesized, or the amount of gene silencing that occurs (see U.S. Patent No. 5,583,021 which is hereby

incorporated by reference). Techniques are known for the in vitro culture of plant tissue, and in a number of cases, for regeneration into whole plants (EP Appln No. 88810309.0). Procedures for transferring the introduced expression complex to commercially useful cultivars are known to those skilled in the art.

Once plant cells expressing the desired level of P450 enzyme are obtained, plant tissues and whole plants can be regenerated therefrom using methods and techniques well-known in the art. The regenerated plants are then reproduced by conventional means and the introduced genes can be transferred to other strains and cultivars by conventional plant breeding techniques.

The following examples illustrate methods for carrying out the invention and should be understood to be illustrative of, but not limiting upon, the scope of the invention which is defined in the appended claims.

EXAMPLES

EXAMPLE I: DEVELOPMENT OF PLANT TISSUE AND ETHYLENE TREATMENT

Plant Growth

Plants were seeded in pots and grown in a greenhouse for 4 weeks. The 4 week old seedlings were transplanted into individual pots and grown in the greenhouse for 2 months. The plants were watered 2 times a day with water containing 150ppm NPK fertilizer during growth. The expanded green leaves were detached from plants to do the ethylene treatment described below.

Cell Line 78379

Tobacco line 78379, which is a burley line released by the University of Kentucky was used as a source of plant material. One hundred plants were cultured as standard in the art of growing tobacco and transplanted and tagged with a distinctive number (1-100). Fertilization and field management were conducted as recommended.

Three quarters of the 100 plants converted between 20 and 100% of the nicotine to nornicotine. One quarter of the 100 plants converted less than 5% of the nicotine to nornicotine. Plant number 87 had the least conversion (2%) while plant number 21 had 100% conversion. Plants converting less than 3% were classified as non-converters. Self-pollinated seed of plant number 87 and plant number 21, as well as crossed (21 x 87 and 87 x 21) seeds were made to study genetic and phenotype differences. Plants from selfed 21 were converters, and 99% of selfs from 87 were non-converters. The other 1% of the plants from 87 showed low conversion (5-15%). Plants from reciprocal crosses were all converters.

Cell Line 4407

Nicotiana line 4407, which is a burley line was used as a source of plant material. Uniform and representative plants (100) were selected and tagged. Of the 100 plants 97 were non-converters and three were converters. Plant number 56 had the least amount of conversion (1.2%) and plant number 58 had the highest level of conversion (96%). Self-pollenated seeds and crossed seeds were made with these two plants.

Plants from selfed-58 were segregating in about a 3:1 converter to non-converter ratio. The 58-33 and 58-25 were identified as homozygous converter and nonconverter plant

lines, respectively. The stable conversion of 58-33 was confirmed by analysis of its progenies of next generation.

Ethylene Treatment Procedures

Green leaves were detached from 2-3 month greenhouse grown plants and sprayed with 0.3% ethylene solution (Prep brand Ethephon (Rhone-Poulenc)). Each sprayed leaf was hung in a curing rack equipped with humidifier and covered with plastic. During the treatment, the sample leaves were periodically sprayed with the ethylene solution. Approximately 24-48 hour post ethylene treatment, leaves were collected for RNA extraction. Another sub-sample was taken for metabolic consituents analysis to determine the concentration of leaf metabolites and more specific constituents of interest such as a variety of alkaloids.

As an example, alkaloids analysis could be performed as follows. Samples (0.1 g) were shaken at 150 rpm with 0.5 ml 2N NaOH, and a 5 ml extraction solution which contained quinoline as an internal standard and methyl t-butyl ether. Samples were analyzed on a HP 6890 GC equipped with a FID detector. A temperature of 250°C was used for the detector and injector. An HP column (30m-0.32nm-1·m) consisting of fused silica crosslinked with 5% phenol and 95% methyl silicon was used at a temperature gradient of 110-185 °C at 10°C per minute. The column was operated at a flow rate at 100°C at 1.7cm^3min^{-1} with a split ratio of 40:1 with a 2·1 injection volume using helium as the carrier gas.

EXAMPLE 2: RNA ISOLATION

For RNA extractions, middle leaves from 2 month old greenhouse grown plants were treated with ethylene as described. The 0 and 24-48 hours samples were used for RNA

extraction. In some cases, leaf samples under the senescence process were taken from the plants 10 days post flower-head removal. These samples were also used for extraction. Total RNA was isolated using Rneasy Plant Mini Kit (Qiagen, Inc., Valencia, California) following manufacturer's protocol.

The tissue sample was grinded under liquid nitrogen to a fine powder using a DEPC treated mortar and pestle. Approximately 100 mg of ground tissue was transferred to a sterile 1.5 ml eppendorf tube. This sample tube was placed in liquid nitrogen until all samples were collected. Then, 450µl of Buffer RLT as provided in the kit (with the addition of β-Mercaptoethanol) was added to each individual tube. The sample was vortexed vigorously and incubated at 56° C for 3 minutes. The lysate was then, applied to the QIAshredder spin column sitting in a 2-ml collection tube, and centrifuged for 2 minutes at maximum speed. The flow through was collected and 0.5 volume of ethanol was added to the cleared lysate. The sample is mixed well and transferred to an Rneasy mini spin column sitting in a 2 ml collection tube. The sample was centrifuged for 1 minute at 10,000rpm. Next, 700µl of buffer RW1 was pipeted onto the Rneasy column and centrifuged for 1 minute at 10,000rpm. Buffer RPE was pipetted onto the Rneasy column in a new collection tube and centrifuged for 1 minute at 10,000 rpm. Buffer RPE was again, added to the Rneasy spin column and centrifuged for 2 minutes at maximum speed to dry the membrane. To eliminate any ethanol carry over, the memebrane was placed in a separate collection tube and centrifuged for an additional 1 minute at maximum speed. The Rneasy column was transfered into a new 1.5 ml collection tube, and 40 µl of Rnase-free water was pipetted directly onto the Rneasy membrane. This final elute tube was centrifuged for 1 minute at 10,000rpm. Quality and quantity of total RNA was analyzed by denatured formaldehyde gel and spectrophotometer.

Poly(A)RNA was isolated using Oligotex poly A RNA purification kit (Qiagen Inc.) following manufacture's protocol. About 200 µg total RNA in 250 µl maximum volume was used. A volume of 250µl of Buffer OBB and 15 µl of Oligotex suspension was added to the 250 µl of total RNA. The contents were mixed thoroughly by pipetting and incubated for 3 minutes at 70°C on a heating block. The sample was then, placed at room temperature for approximately 20 minutes. The oligotex:mRNA complex was pelleted by centrifugation for 2 minutes at maximum speed. All but 50 µl of the supernatant was removed from the microcentrifuge tube. The Sample was treated further by OBB buffer. The oligotex:mRNA pellet was resuspended in 400 µl of Buffer OW2 by vortexing. This mix was transferred onto a small spin column placed in a new tube and centrifuged for 1 minute at maximum speed. The spin column was transferred to a new tube and an additional 400 µl of Buffer OW2 was added to the column. The tube was then centrifuged for 1 minute at maximum speed. The spin column was transferred to a final 1.5ml microcentrifuge tube. The sample was eluted with 60 ul of hot (70 C) Buffer OEB. Poly A product was analyzed by denatured formaldehyde gels and spectrophotometric analysis.

EXAMPLE 3: REVERSE TRANSCRIPTION-PCR

First strand cDNA was produced using SuperScript reverse transcriptase following manufacturer's protocol(Invitrogen, Carlsbad, California). The poly A enriched RNA/oligo dT primer mix consisted of less than 5 µg of total RNA, 1 µl of 10mM dNTP mix, 1 µl of Oligo $d(T)_{12-18}$ (0.5µg/µl), and up to 10 µl of DEPC-treated water. Each sample was incubated at 65° C for 5 minutes, then placed on ice for at least 1 minute. A reaction mixture was prepared by adding each of the following components in order: 2 µl 10X RT buffer, 4 µl of 25 mM MgCl2, 2µl of 0.1 M DTT, and 1 µl of RNase OUT Recombinant RNase

Inhibitor. An addition of 9 µl of reaction mixture was pipetted to each RNA/primer mixture and gently mixed. It was incubated at 42° C for 2 minutes and 1 µl of Super Script II RT was added to each tube. The tube was incubated for 50 minutes at 42° C. The reaction was terminated at 70° C for 15 minutes and chilled on ice. The sample was collected by centrifugation and 1 µl of RNase H was added to each tube and incubated for 20 minutes at 37° C. The second PCR was carried out with 200 pmoles of forward primer (degenerate primers as in Figure 75, SEQ.ID Nos. 149-156) and 100 pmoles reverse primer (mix of 18nt oligo d(T) followed by 1 random base).

Reaction conditions were 94°C for 2 minutes and then performed 40 cycles of PCR at 94°C for 1 minute, 45° to 60°C for 2 minutes, 72°C for 3 minutes with a 72°C extension for an extra 10 min.

Ten microliters of the amplified sample were analyzed by electrophoresis using a 1% agarose gel. The correct size fragments were purified from agarose gel.

EXAMPLE 4: GENERATION OF PCR FRAGMENT POPULATIONS

PCR fragments from Example 3 were ligated into a pGEM-T Easy Vector (Promega, Madison, Wisconsin) following manufacturer's instructions. The ligated product was transformed into JM109 competent cells and plated on LB media plates for blue/white selection. Colonies were selected and grown in a 96 well plate with 1.2 ml of LB media overnight at 37°C. Frozen stock was generated for all selected colonies. Plasmid DNA from plates were purified using Beckman's Biomeck 2000 miniprep robotics with Wizard SV Miniprep kit (Promega). Plasmid DNA was eluted with 100µl water and stored in a 96 well plate. Plasmids were digested by EcoR1 and were analyzed using 1% agarose gel to confirm the DNA quantity and size of inserts. The plasmids containing a 400-600 bp insert were sequenced using an CEQ 2000 sequencer (Beckman, Fullerton, California). The sequences were

aligned with GenBank database by BLAST search. The P-450 related fragments were identified and further analyzed.

EXAMPLE 5: CONSTRUCTION OF CDNA LIBRARY

A cDNA library was constructed by preparing total RNA from ethylene treated leaves as follows. First, total RNA was extracted from ethylene treated leaves of tobacco line 58-33 using a modified acid phenol and chloroform extraction protocol. Protocol was modified to use one gram of tissue that was ground and subsequently vortexed in 5 ml of extraction buffer (100 mM Tris-HCl, pH 8.5; 200 mM NaCl; 10mM EDTA; 0.5% SDS) to which 5 ml phenol (pH5.5) and 5 ml chloroform was added. The extracted sample was centrifuged and the supernatant was saved. This extraction step was repeated 2-3 more times until the supernatant appeared clear. Approximately 5 ml of chloroform was added to remove trace amounts of phenol. RNA was precipitated from the combined supernatant fractions by adding a 3-fold volume of ETOH and 1/10 volume of 3M NaOAc (pH5.2) and storing at -20° C for 1 hour. After transferring to Corex glass container it was centrifuged at 9,000 RPM for 45 minutes at 4° C. The pellet was washed with 70% ethanol and spun for 5 minutes at 9,000 RPM at 4° C. After drying the pellet, the pelleted RNA was dissolved in 0.5 ml RNase free water. The pelleted RNA was dissolved in 0.5 ml RNase free water. The quality and quantity of total RNA was analyzed by denatured formaldehyde gel and spectrophotometer, respectively.

The resultant total RNA was isolated for poly A+ RNA using an Oligo(dT) cellulose protocol (Invitrogen) and Microcentrifuge spin columns (Invitrogen) by the following protocol. Approximately twenty mg of total RNA was subjected to twice purification to obtain high quality poly A+ RNA. Poly A+ RNA product was analyzed by performing denatured formaldehyde gel and subsequent RT-PCR of known full-length genes to ensure high quality of mRNA. In addition, Northern analysis was performed on the poly A+RNA from ethylene treated non-converter

leaves, zero hour ethylene treated converter leaves and ethylene treated converter leaves using the full-length p450 as probe. The method was based on the protocol provided by the manufacturer's instructions (KPL RNADetector Northern Blotting Kit, Gaithersburg, Maryland) using 1.8 μg of polyA+RNA for each sample. RNA containing gels were transferred overnight using 20X SSC as a transfer buffer.

Next, poly A+ RNA was used as template to produce a cDNA library employing cDNA synthesis kit, ZAP-cDNA synthesis kit, and ZAP-cDNA Gigapack III gold cloning kit (Stratagene, La Jolla, California). The method involved following the manufacture's protocol as specified. Approximately 8 μg of poly A+ RNA was used to construct cDNA library. Analysis of the primary library revealed about $2.5 \times 10^6 - 1 \times 10^7$ pfu. A quality background test of the library was completed by complementation assays using IPTG and X-gal, where recombinant plaques was expressed at more than 100-fold above the background reaction.

A more quantitative analysis of the library by random PCR showed that average size of insert cDNA was approximately 1.2 kb. The method used a two-step PCR method as followed. For the first step, reverse primers were designed based on the preliminary sequence information obtained from P450 fragments. The designed reverse primers and T3 (forward) primers were used amplify corresponding genes from the cDNA library. PCR reactions were subjected to agarose electrophoresis and the corresponding bands of high molecular weight were excised, purified, cloned and sequenced. In the second step, new primers designed from 5'UTR or the start coding region of P450 as the forward primers together with the reverse primers (designed from 3'UTR of P450) were used in the subsequent PCR to obtain full-length P450 clones.

The P450 fragments were generated by PCR amplification from the constructed cDNA library as described in example 3 with the exception of the reverse primer. The T7 primer located on the plasmid downstream of cDNA inserts (see Figure 75), was used as a reverse primer. PCR

fragments were isolated, cloned and sequenced as described in Example 4.

Numerous modifications and variations in practice of the invention are expected to occur to those skilled in the art upon consideration of the foregoing detailed description of the invention. Consequently, such modifications and variations are intended to be included within the scope of the following claims.

EXAMPLE 6: CHARACTERIZATION OF CLONED FRAGMENTS – REVERSE SOUTHERN BLOTTING ANALYSIS

Nonradioactive large scale reverse southern blotting assay was performed on all P450 clones identified in above examples to detect the differential expression. It was observed that the level of expression among different P450 clusters was very different. Further real time detection was conducted on those with high expression.

Nonradioactive southern blotting procedures were conducted as follows.

1) Total RNA was extracted from ethylene treated converter (58-33) and nonconverter (58-25) leaves using the Qiagen Rnaeasy kit as described in Example 2.

2) Probe was produced by biotin-tail labeling a single strand cDNA derived from poly A enriched RNA generated in above step. This labeled single strand cDNA was generated by RT-PCR of the converter and nonconverter total RNA (Invitrogen) as described in example 3 with the exception of using biotinalyted oligo dT as a primer (Promega); These were used as a probe to hybridize with cloned DNA.

3) Plasmid DNA was digested with restriction enzyme EcoR1 and run on agarose gels. Gels were simultaneously dried and transferred to two nylon membranes (Biodyne B). One membrane was hybridized with converter probe and the other with nonconverter probe. Membranes were UV-crosslinked (auto crosslink setting, 254 nm, Stratagene, Stratalinker) before hybridization.

Alternatively, the inserts were PCR amplified from each plasmid using the sequences located on both arms of p-GEM plasmid, T3 and SP6, as primers. The PCR products were analyzed by running on a 96 well Ready-to-run agarose gels. The confirmed inserts were dotted on two nylon membranes. One membrane was hybridized with converter probe and the other with nonconverter probe.

4). The membranes were hybridized and washed following manufacture's instruction with the modification of washing stringency (Enzo Diagnostics, Inc, Farmingdale, NY). The membranes were prehybridized with hybridization buffer (2x SSC buffered formamide, containing detergent and hybridization enhancers) at 42°C for 30 min and hybridized with 10μl denatured probe overnight at 42°C. The membranes then were washed in 1X hybridization wash buffer 1 time at room temperature for 10 min and 4 times at 68°C for 15 min. The membranes were ready for the detection.

5) The washed membranes were detected by alkaline phosphatase labeling followed by NBT/BCIP colometric detection as described in manufacture's detection procedure (Enzo Diagnostics, Inc.). The membranes were blocked for one hour at room temperature with 1x blocking solution, washed 3 times with 1X detection reagents for 10 min, washed 2 times with 1x predevelopment reaction buffer for 5 min and then developed the blots in developing solution for 30-45 min until the dots appear. All reagents were provided by manufacture (Enzo Diagnostics, Inc).

In some cases, one step RT-PCR (Gibco Kit, Carlsbad, California) was performed on the total RNA's from non-converter (58-25) and converter (58-33) lines using primers specific to the P-450 fragments. Comparative RT-PCR was conducted as follows:

1) Total RNA from ethylene treated converter (58-33) and nonconverter (58-25) plant leaves was extracted as described in example 2.

2) Poly(A) RNA from total RNA was extracted using Qiagen kit as described in example 2.

3) One step RT-PCR was conducted using primers specific to cloned P450 following the manufactures procedure(Invitogen). The poly A enriched RNA was added to the reaction mix, along with, 25µl of 2X Reaction Mix, 1µl of 10µM Sense Primer, 1µl of 10 µM Anti-sense Primer, 1 µl of RT/ Platinum taq Mix, and up to 50 µl of water. Reaction conditions were 50°C for 20 minutes and then 94 C for 2 min, performed 40 cycles of PCR at 94°C for 30 sec, 55° to for 30 sec, 70°C for 1 minute with a 72°C extension for an extra 10 min. Ten microliters of the amplified sample were analyzed by electrophoresis using a 1% agarose gel.

EXAMPLE 7: CHARACTERIZATION OF CLONED FRAGMENTS - NORTHERN BLOT ANALYSIS

Alternative to Southern Blot analysis, some membranes were hybridized and detected as described in the example of northern blotting assays. Northern Hybridization was used to detect mRNA differentially expressed in Nicotiana as follows.

First step, probe preparation: the random priming method was used to prepare probes from cloned p450 DNA fragments (Random Primer DNA Biotinylation Kit, KPL). The following components were mixed: 0.5µg DNA template (boiled in a water bath for 5-10 minutes and chilled on ice before use); 1X Random Primer Solution; 1X dNTP mix; 10 units of Klenow and water was added to bring the reaction to 50µl. The mixture was incubated in 37 °C for 1-4 hours. The reaction was stopped with 2µl of 200 mM EDTA. The probe was denatured by incubating at 95 °C for 5 minutes before use.

Second step, sample preparation: The RNA samples were prepared from ethylene treated and non-treated fresh leaves, and senescence

leaves. In some cases poly A enriched RNA was used. Approximately 15μg total RNA or 1.8μg mRNA (Methods of RNA and mRNA extraction are described in Example 5) was brought to equal volume with DEPC H2O(5-10 μl). The same volume loading buffer (1 x MOPS; 18.5 % Formaldehyde; 50 % Formamide; 4 % Ficoll400; Bromophenolblue) and 0.5 μl EtBr (0.5μg /μl) were added. The samples were heated at 90 °C for 5 minutes, and chilled on ice.

Third step, separation of RNA by electrophoresis: Samples were subjected to electrophoresis on a formaldehyde gel (1 % Agarose, 1 x MOPS, 0.6 M Formaldehyde) with 1XMOP buffer (0.4 M Morpholinopropanesulfonic acid; 0.1 M Na-acetate-3 x H2O; 10 mM EDTA; adjust to pH 7.2 with NaOH). RNAs were transferred to Hybond-N+ membrane (Nylon, Amersham Pharmacia Biotech) by capillary method in 10 X SSC buffer (1.5 M NaCl; 0.15 M Na-citrate) for 24 hours. Membranes with RNA samples were UV-crosslinked (auto crosslink setting, 254 nm, Stratagene, Stratalinker) before hybridization.

Fourth step, hybridization: The membrane was prehybridized for 1-4 hours at 42 °C with 5-10 ml prehybridization buffer (5 x SSC; 50 % Formamide; 5 x Denhardt's-solution; 1 % SDS; 100μg/ml heat-denatured sheared non- homologous DNA). Old prehybridization buffer was discarded, and new prehybridization buffer and probe were added. The hybridization was carried out over night at 42 °C. The membrane was washed for 15 minutes with 2 x SSC at room temperature, followed by a wash with 2 x SSC, 0.1 % SDS at 65 °C for 2 times, and a final wash with 0.1 x SSC, or more wash with 0.1 x SDS at 65 °C (optional).

Fifth step, detection: AP-Streptavidin and CDP-Star were used to detect the hybridization signal(KPL's DNA Detector Northern blotting Kit). The membrane was blocked with 1X Detector Block Solution for 30 minutes at room temperature. The blocking buffer was discarded and the membrane was incubated in new 1X detector Block Solution with 1:10,000 AP-SA at room temperature for 1 hour. The membrane was washed in 1X Phosphatase Wash Solution for 3 times, followed by a wash with 1X Phosphatase Assay Buffer for two times. The signal was detected with CDP-Star Chemiluminescent Substrate. The wet membrane was exposed to X-Ray film under saran™ wrap. The results were analyzed and recorded.

A major focus of the invention was the discovery of novel genes that may be induced as a result of ethylene treatment or play a key role in tobacco leaf quality and constituents. As shown in the table below, Northern blots were useful in determining which genes were induced by ethylene treatment relative to non-induced plants. Interestingly, not all fragments were affected similarly in the converter and nonconverter. The cytochrome P450 fragments of interest were partially sequenced to determine their structural relatedness. This information was used to subsequently isolate and sequence full length gene clones. Functional analysis utilizing down-regulation methods was performed in whole plants with the fragments genes.

Fragments	Induced mRNA Expression Ethylene Treatment	
	Converter	Nonconverter
D186-AH4	+	
D56-AC7	+	+
D56-AG11	+	
D56-AC12	+	+
D70A-AB5	+	+
D73-AC9	+	+
D70A-AA12	+	+
D73A-AG3	+	
D73A-AE10		+
D35-AG11	+	
D58-AD4	+	+
D34-52	+	+
D56-AG6	+	+

EXAMPLE 8: NUCLEIC ACID IDENTITY AND STRUCTURE RELATEDNESS OF ISOLATED NUCLEIC ACID FRAGMENTS

Over 100 cloned P450 fragments were sequenced in conjunction with Northern blot analysis to determine their structural relatedness. The approach used utilized forward primers based either of two common P450 motifs located near the carboxyl-terminus of the P450 genes. The forward primers corresponded to cytochrome P450 motifs FXPERF or GRRXCP(A/G) as denoted in Figure 1. The reverse primers used standard primers from either the plasmid, SP6 or T7 located on both arms of pGEM plasmid, or a poly A tail. The protocol used is described below.

Spectrophotometry was used to estimate the concentration of starting double stranded DNA following the manufacturer's protocol (Beckman Coulter). The template was diluted with water to the appropriate concentration, denatured by heating at 95° C for 2 minutes, and subsequently placed on ice. The sequencing reaction was prepared on ice using 0.5 to 10µl of denatured DNA template, 2 µl of 1.6 pmole of the forward primer, 8 µl of DTCS Quick Start Master Mix and the total volume brought to 20 µl with water. The thermocycling

program consisted of 30 cycles of the follow cycle: 96° C for 20 seconds, 50° C for 20 seconds, and 60° C for 4 minutes followed by holding at 4° C.

The sequence was stopped by adding 5 µl of stop buffer (equal volume of 3M NaOAc and 100mM EDTA and 1 µl of 20 mg/ml glycogen). The sample was precipitated with 60 µl of cold 95% ethanol and centrifuged at 6000g for 6 minutes. Ethanol was discarded. The pellet was 2 washes with 200 µl of cold 70% ethanol. After the pellet was dry, 40 µl of SLS solution was added and the pellet was resuspended. A layer of mineral oil was over laid. The sample was then, placed on the CEQ 8000 Automated Sequencer for further analysis.

In order to verify nucleic acid sequences, nucleic acid sequence was re-sequenced in both directions using forward primers to the FXPERF or GRRXCP(A/G) region of the P450 gene or reverse primers to either the plasmid or poly A tail. All sequencing was performed at least twice in both directions.

The nucleic acid sequences of cytochrome P450 fragments were compared to each other from the coding region corresponding to the first nucleic acid after the region encoding the GRRXCP(A/G) motif through to the stop codon. This region was selected as an indicator of genetic diversity among P450 proteins. A large number of genetically distinct P450 genes, in excess of 70 genes, was observed similar to that of other plant species. Upon comparison of nucleic acid sequences, it was found that the genes could be placed into distinct sequences groups based on their sequence identity. It was found that the best unique grouping of P450 members was determined to be those sequences with 75% nucleic acid identity or greater (shown in Table I). Reducing the percentage identity resulted in significantly larger groups. A preferred grouping was observed for those sequences with 81% nucleic acid identity or greater, a more preferred grouping 91% nucleic acid identity or greater, and a most preferred grouping for those sequences 99% nucleic acid identity of greater. Most of the

groups contained at least two members and frequently three or more members. Others were not repeatedly discovered suggesting that approach taken was able to isolated both low and high expressing mRNA in the tissue used.

Based on 75% nucleic acid identity or greater, two cytochrome P450 groups were found to contain nucleic acid sequence identity to previously tobacco cytochrome genes that genetically distinct from that within the group. Group 23, showed nucleic acid identity, within the parameters used for Table I, to prior GenBank sequences of GI:1171579 (CAA64635) and GI:14423327 (or AAK62346) by Czernic et al and Ralston et al, respectively. GI:1171579 had nucleic acid identity to Group 23 members ranging 96.9% to 99.5% identity to members of Group 23 while GI:14423327 ranged 95.4% to 96.9% identity to this group. The members of Group 31 had nucleic acid identity ranging from 76.7% to 97.8% identity to the GenBank reported sequence of GI:14423319 (AAK62342) by Ralston et al. None of the other P450 identity groups of Table 1 contained parameter identity, as used in Table 1, to Nicotiana P450s genes reported by Ralston et al, Czernic et al., Wang et al or LaRosa and Smigocki.

As shown in Figure 76, consensus sequence with appropriate nucleic acid degenerate probes could be derived for group to preferentially identify and isolate additional members of each group from Nicotiana plants.

Table I: Nicotiana P450 Nucleic Acid Sequence Identity Groups

GROUP	FRAGMENTS
1	D58–BG7 (SEQ ID No.:1), D58-AB1 (SEQ ID No.:3); D58-BE4 (SEQ ID No.:7)
2	D56-AH7 (SEQ ID No.:9); D13a-5 (SEQ ID No.:11)
3	D56-AG10 (SEQ ID No.:13); D35-33 (SEQ ID No.:15); D34-62) (SEQ ID No.:17)
4	D56-AA7 (SEQ ID No.:19); D56-AE1 (SEQ ID No.:21); 185-BD3 (SEQ ID No.:143)
5	D35-BB7 (SEQ ID No.:23); D177-BA7 (SEQ ID No.:25) D56A-AB6 (SEQ ID No.:27); D144-AE2 (SEQ ID No.:29)
6	D56-AG11 (SEQ ID No.:31); D179-AA1 (SEQ ID No.:33)
7	D56-AC7 (SEQ ID No.:35); D144-AD1 (SEQ ID No.:37)
8	D144-AB5 (SEQ ID No.:39)
9	D181-AB5 (SEQ ID No.:41); D73-Ac9 (SEQ ID No.:43)
10	D56-AC12 (SEQ ID No.:45)
11	D58-AB9 (SEQ ID No.:47); D56-AG9 (SEQ ID No.:49); D56-AG6 (SEQ ID No.:51); D35-BG11 (SEQ ID No.:53); D35-42 (SEQ ID No.:55); D35-BA3 (SEQ ID No.:57); D34-57 (SEQ ID No.:59); D34-52 (SEQ ID No.:61); D34-25 (SEQ ID No.:63)
12	D56-AD10 (SEQ ID No.:65)
13	56-AA11 (SEQ ID No.:67)
14	D177-BD5 (SEQ ID No.:69); D177-BD7 (SEQ ID No.:83)
15	D56A-AG10 (SEQ ID No.:71); D58-BC5 (SEQ ID No.:73); D58-AD12 (SEQ ID No.:75)
16	D56-AC11 (SEQ ID No.:77); D35-39 (SEQ ID No.:79); D58-BH4 (SEQ ID No.:81); D56-AD6 (SEQ ID No.:87)

17 D73A-AD6 (SEQ ID No.:89); D70A-BA11 (SEQ ID No.:91); D70A-BB5 (SEQ ID No.:93)

18 D70A-AB5 (SEQ ID No.:95); D70A-AA8 (SEQ ID No.:97)

19 D70A-AB8 (SEQ ID No.:99); D70A-BH2 (SEQ ID No.:101); D70A-AA4 (SEQ ID No.:103)

20 D70A-BA1 (SEQ ID No.:105); D70A-BA9 (SEQ ID No.:107); D176-BG2 (SEQ ID No.:141)

21 D70A-BD4 (SEQ ID No.:109)

22 D181-AC5 (SEQ ID No.:111); D144-AH1 (SEQ ID No.:113); D34-65 (SEQ ID No.:115)

23 D35-BG2 (SEQ ID No.:117)

24 D73A-AH7 (SEQ ID No.:119)

25 D58-AA1 (SEQ ID No.:121); D185-BC1 (SEQ ID No.:133); D185-BG2 (SEQ ID No.:135)

26 D73-AE10 (SEQ ID No.:123)

27 D56-AC12 (SEQ ID No.:125)

28 D177-BF7 (SEQ ID No.:127); D185-BE1 (SEQ ID No.:137); 185-BD2 (SEQ ID No.:139)

29 D73A-AG3 (SEQ ID No.:129)

30 D70A-AA12 (SEQ ID No.:131); D176-BF2 (SEQ ID No.:85)

31 D176-BC3 (SEQ ID No.:145)

32 D176-BB3 (SEQ ID No.: 147)

33 D186-AH4 (SEQ ID No.:5)

EXAMPLE 9: RELATED AMINO ACID SEQUENCE IDENTITY OF ISOLATED NUCLEIC ACID FRAGMENTS

The amino acid sequences of nucleic acid sequences obtained for cytochrome P450 fragments from Example 8 were deduced. The deduced region corresponded to the amino acid immediately after the

GXRXCP(A/G) sequence motif to the end of the carboxyl-terminus, or stop codon. Upon comparison of sequence identity of the fragments, a unique grouping was observed for those sequences with 70% amino acid identity or greater. A preferred grouping was observed for those sequences with 80% amino acid identity or greater, more preferred with 90% amino acid identity or greater, and a most preferred grouping for those sequences 99% amino acid identity of greater. The groups and corresponding amino acid sequences of group members are shown in Figure 2. Several of the unique nucleic acid sequences were found to have complete amino acid identity to other fragments and therefore only one member with the identical amino acid was reported.

The amino acid identity for Group 19 of Table II corresponded to three distinct groups based on their nucleic acid sequences. The amino acid sequences of each group member and their identity is shown in Figure. 77. The amino acid differences are appropriated marked.

At least one member of each amino acid identity group was selected for gene cloning and functional studies using plants. In addition, group members that are differentially affected by ethylene treatment or other biological differences as assessed by Northern and Southern analysis were selected for gene cloning and functional studies. To assist in gene cloning, expression studies and whole plant evaluations, peptide specific antibodies will be prepared on sequence identity and differential sequence.

Table II: Nicotiana P450 Amino Acid Sequence Identity Groups

GROUP	FRAGMENTS
1	D58—BG7 (SEQ ID No.:2), D58-AB1 (SEQ ID No.:4)
2	D58-BE4 (SEQ ID No.:8)
3	D56-AH7 (SEQ ID No.:10); D13a-5 (SEQ ID No.:12)

4 D56-AG10 (SEQ ID No.:14); D34-62 (SEQ ID No.:18)

5 D56-AA7 (SEQ ID No.:20); D56-AE1 (SEQ ID No.:22); 185-BD3 (SEQ ID No.:144)

6 D35-BB7 (SEQ ID No.:24); D177-BA7 (SEQ ID No.:26); D56A-AB6 (SEQ ID No.:28); D144-AE2 (SEQ ID No.:30)

7 D56-AG11 (SEQ ID No.:32); D179-AA1 (SEQ ID No.:34)

8 D56-AC7 (SEQ ID No.:36); D144-AD1 (SEQ ID No.:38)

9 D144-AB5 (SEQ ID No.:40)

10 D181-AB5 (SEQ ID No.:42); D73-Ac9 (SEQ ID No.:44)

11 D56-AC12 (SEQ ID No.:46)

12 D58-AB9 (SEQ ID No.:48); D56-AG9 (SEQ ID No.:50); D56-AG6 (SEQ ID No.:52); D35-BG11 (SEQ ID No.:54); D35-42 (SEQ ID No.:56); D35-BA3 (SEQ ID No.:58); D34-57 (SEQ ID No.:60); D34-52 (SEQ ID No.:62)

13 D56AD10 (SEQ ID No.:66)

14 56-AA11 (SEQ ID No.:68)

15 D177-BD5 (SEQ ID No.:70); D177-BD7 (SEQ ID No.:84)

16 D56A-AG10 (SEQ ID No.:72); D58-BC5 (SEQ ID No.:74); D58-AD12 (SEQ ID No.:76)

17 D56-AC11 (SEQ ID No.:78); D56-AD6 (SEQ ID No.:88)

18 D73A-AD6 (SEQ ID No.90:); D70A-BB5 (SEQ ID No.:94)

19 D70A-AB5 (SEQ ID No.:96); D70A-AB8 (SEQ ID No.:100); D70A-BH2 (SEQ ID No.:102); D70A-AA4 (SEQ ID No.:104); D70A-BA1 (SEQ ID No.:106); D70A-BA9 (SEQ ID No.:108); D176-BG2 (SEQ ID No.:142)

20 D70A-BD4 (SEQ ID No.:110)

21 D181-AC5 (SEQ ID No.:112); D144-AH1 (SEQ ID No.:114); D34-65 (SEQ ID No.:116)

22 D35-BG2 (SEQ ID No.:118)

23 D73A-AH7 (SEQ ID No.:120)

24 D58-AA1 (SEQ ID No.:122); D185-BC1 (SEQ ID No.:134); D185-BG2 (SEQ ID No.:136)

25 D73-AE10 (SEQ ID No.:124)

26 D56-AC12 (SEQ ID No.:126)

27 D177-BF7 (SEQ ID No.:128); 185-BD2 (SEQ ID No.:140)

28 D73A-AG3 (SEQ ID No.:130)

29 D70A-AA12 (SEQ ID No.:132); D176-BF2 (SEQ ID No.:86)

30 D176-BC3 (SEQ ID No.:146)

31 D176-BB3 (SEQ ID No.:148)

32 D186-AH4 (SEQ ID No.:6)

EXAMPLE 10: CLONING OF FULL LENGTH cDNA P450 CLONES

A cDNA library was constructed by preparing total RNA from ethylene treated leaves as follows. First, total RNA was extracted from ethylene treated leaves using a modified acid phenol and chloroform extraction protocol. Protocol was modified to use one gram of tissue that was ground and subsequently vortexed in 5 ml of extraction buffer (100 mM Tris-HCl, pH 8.5; 200 mM NaCl; 10mM EDTA; 0.5% SDS) to which 5 ml phenol (pH5.5) and 5 ml chloroform was added. The extracted sample was centrifuged and the supernatant was saved. This extraction step was repeated 2-3 more times until the supernatant appeared clear. Approximately 5 ml of chloroform was added to remove trace amounts of phenol. RNA was precipitated from the combined supernatant fractions by adding a 3-fold volume of ETOH and 1/10 volume of 3M NaOAc (pH5.2) and storing at -20° C for 1 hour. After transfering to Corex glass container it was centrifuged at 9,000 RPM for 45 minutes at 4°

C. The pellet was washed with 70% ethanol and spun for 5 minutes at 9,000 RPM at 4° C. After drying the pellet, the pelleted RNA was dissolved in 0.5 ml RNase free water. The pelleted RNA was dissolved in 0.5 ml RNase free water. The quality and quantity of total RNA was analyzed by denatured formaldehyde gel and spectrophotometer, respectively.

The resultant total RNA was isolated for poly A+ RNA using an Oligo(dT) cellulose protocol (Invitrogen) and Microcentrifuge spin columns (Invitrogene) by the following protocol. Approximately twenty mg of total RNA was subjected to twice purification to obtain high quality poly A+ RNA. Poly A+ RNA product was analyzed by performing denatured formaldehyde gel and subsequent RT-PCR of known full-length genes to ensure high quality of mRNA. In addition, Northern analysis was performed on the poly A+RNA from ethylene treated non-converter leaves, zero hour ethylene treated converter leaves and ethylene treated converter leaves using the full-length p450 as probe. The method was based on the protocol provided by the manufacturer's instructions (KPL RNADetector Northern Blotting Kit) using 1.8 ug of polyA+RNA for each sample. RNA containing gels were transferred overnight using 20X SSC as a transfer buffer.

Next, poly A+ RNA was used as template to produce a cDNA library employing cDNA synthesis kit, ZAP-cDNA synthesis kit, and ZAP-cDNA Gigapack III gold cloning kit (Stratagene). The method involved following the manufacture's protocol as specified. Approximately 8 ug of poly A+ RNA was used to construct cDNA library. Analysis of the primary library revealed about 2.5 x 106 - 1x 107 pfu. A quality background test of the library was completed by a- complementation using IPTG and X-gal, where recombinant plaques was expressed at more than 100-fold above the background reaction.

A more quantitative analysis of the library by random PCR showed that average size of insert cDNA was approximately 1.2 kb. The method used a two-step PCR method as followed. For the first step, reverse primers were designed based on the preliminary sequence information obtained from p450 fragments. The designed reverse primers and T3 (forward) primers were used amplify corresponding genes from the cDNA library. PCR reactions were subjected to agarose electrophoresis and the corresponding bands of high molecular weight were excised, purified, cloned and sequenced. In the second step, new primers designed from 5'UTR or the start coding region of p450s as the forward primers together with the reverse primers (designed from 3'UTR of p450) were used in the subsequent PCR to obtain full-length p450 clones.

Full-length p450 genes were isolated by PCR method from constructed cDNA library. Two steps of PCR were used to clone the full-length genes. In the first step PCR, unspecific reverse primer (T3) and specific forward primer (generated from the downstream sequence of P450s) were used to clone the 5'end of the P450s from cDNA library. PCR fragments were isolated, cloned and sequenced for designing the forward primers in next step PCR. Two specific primers were used to clone the full-length p450 clones in the second step PCR. The clones were subsequently sequenced.

Numerous modifications and variations in practice of the invention are expected to occur to those skilled in the art upon consideration of the foregoing detailed description of the invention. Consequently, such modifications and variations are intended to be included within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleic acid sequence selected from the group consisting of SEQ. ID. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147 or 149.

2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule comprises a fragment of a cytochrome P450 gene.

3. An isolated nucleic acid molecule, wherein said nucleic acid molecule has at least 75% identity to the nucleic acid molecule of Claim 1.

4. An isolated nucleic acid molecule, wherein said nucleic acid molecule has at least 91% identity to the nucleic acid molecule of Claim 1.

5. An isolated nucleic acid molecule, wherein said nucleic acid molecule has at least 99% identity to the nucleic acid molecule of Claim 1.

6. A transgenic plant, wherein said transgenic plant comprises the nucleic acid molecule of claim 1, 2, 3, 4 or 5.

7. The transgenic plant of Claim 6, wherein said plant is a tobacco plant.

8. A method of producing a transgenic plant, said method comprising the steps of:

(i) operably linking the nucleic acid molecule of Claims 1, 2, 3, 4 or 5, with a promoter functional in said plant to create a plant transformation vector; and

(ii) transforming said plant with said plant transformation vector of step (i);

(iii) selecting a plant cell transformed with said transformation vector; and

(iv) regenerating a plant from said plant cell of step (iii).

9. The method of Claim 8, wherein said nucleic acid molecule is in an antisense orientation.

10. The method of Claim 8, wherein said nucleic acid molecule is in a sense orientation.

11. The method of Claim 8 wherein said nucleic acid is in a RNA interference orientation.

12. The method of Claim 11, wherein said nucleic acid molecule is expressed as a double stranded RNA molecule.

13. The method of Claim 11, wherein said double stranded RNA molecule is about 15 to 25 nucleotide in length.

14. The method of Claim 8, wherein said plant is a tobacco plant.

15. A method of selecting a plant containing a nucleic acid molecule, wherein said plant is analyzed for the presence of nucleic acid sequence, wherein said nucleic sequence acid selected from the group consisting of SEQ. ID 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 50, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 92, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145 or 147.

16. The method of selecting a plant of claim 15, wherein said plant is analyzed by DNA hybridization.

17. The method of selecting a plant of claim 15, wherein said plant is analyzed by PCR detection.

18. The method of claim 16, wherein said DNA hybridization comprises a nucleic acid probe, said nucleic acid probe is selected from a group consisting of SEQ. ID. 1, 3, 5, 7, 9, 11, 13, 15, 17,19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145 or 147.

19. The method of selecting a plant of claim 15, wherein said plant is a transgenic plant.

20. The method of selecting a plant of claim 15, wherein said plant is selected from a mutagenesis population.

21. The method of selecting a plant of claim 15, wherein said plant is selected from a breeding population.

FIG. 1

SEQ ID 1 D58-BG7

```
   1 GCACAACTT GCTATCAACT TGGTCACATC TATGTTGGGT
  61 CATTTGTTGC ATCATTTTAC ATGGGCTCCG GCCCCGGGGG TTAACCCGGA GGATATTGAC
 121 TTGGAGGAGA GCCCTGGAAC AGTAACTTAC ATGAAAAATC CAATACAAGC TATTCCAACT
 181 CCAAGATTGC CTGCACACTT GTATGGACGT GTGCCAGTGG ATATGTAA
```

SEQ ID 2

AQLAINLVTSMLGHLLHHFTWAPAPGVNPEDIDLEESPGTVTYMKNPIQAIPTPRLPAHLYGRVPVDM

FIG. 2

SEQ ID 3 D58-AB1

```
   1 GCACAACT TGCTATCAAC TTGGTCACAT CTATGTTGGG
  61 TCATTTGTTG CATCATTTTA CGTGGGCTCC GCCCCCGGGG GTTAACCCGG AGAATATTGA
 121 CTTGGAGGAG AGCCCTGGAA CAGTAACTTA CATGAAAAAT CCAATACAAG CTATTCCTAC
 181 TCCAAGATTG CCTGCACACT TGTATGGACG TGTGCCAGTG GATATGTAA
```

SEQ ID 4

AQLAINLVTSMLGHLLHHFTWAPPPGVNPENIDLEESPGTVTYMKNPIQAIPTPRLPAHLYGRVPVDM

FIG. 3

SEQ ID 5 D186-AH4

```
   1 ATGAATTAT TCATTGCAAG TGGAACACCT TTCAATTGCT
  61 CATATGATCC AAGGTTTCAG TTTTGCAACT ACGACCAATG AGCCTTTGGA TATGAAACAA
 121 GGTGTGGGTT TAACTTTACC AAAGAAGACT GATGTTGAAG TGCTAATTAC ACCTCGCCTT
 181 CCTCCTACGC TTTATCAATA TTAA
```

SEQ ID 6

MNYSLQVEHLSIAHMIQGFSFATTTNEPLDMKQGVGLTLPKKTDVEVLITPRLPPTLYQY

FIG. 4

SEQ ID 7 D58-BE4

```
   1 GCACAACTT GCTATCAACT TGGTCACATC TATGTTGGGT
  61 CATTTGTTCA TCATTTTACA TGGGCTCCGG CCCCGGGGGT TAACCCGGAG GATATTGACT
 121 TGGAGGAGAG CCCTGGAACA GTAACTTACA TGA
```

SEQ ID 8

AQLAINLVTSMLGHLFIILHGLRPRGLTRRILTWRRALEQ

FIG. 5

SEQ ID 9 D56-AH7

```
   1 GAAGGATTG GCTGTTCGAA TGGTTGCCTT GTCATTGGGA
  61 TGTATTATTC AATGTTTTGA TTGGCAACGA ATCGGCGAAG AATTGGTTGA TATGACTGAA
 121 GGAACTGGAC TTACTTTGCC TAAAGCTCAA CCTTTGGTGG CCAAGTGTAG CCCACGACCT
 181 AAAATGGCTA ATCTTCTCTC TCAGATTTGA
```

SEQ ID 10

EGLAVRMVALSLGCIIQCFDWQRIGEELVDMTEGTGLTLPKAQPLVAKCSPRPKMANLLSQI

FIG. 6

SEQ ID 11 D13a-5

```
   1 GAAGGATTG GCTATTCGAA TGGTTGCATT GTCATTGGGA
  61 TGTATTATTC AATGCTTTGA TTGGCAACGA CTTGGGGAAG GATTGGTTGA TAAGACTGAA
 121 GGAACTGGAC TTACTTTGCC TAAAGCTCAA CCTTTAGTGG CCAAGTGTAG CCCACGACCT
 181 ATAATGGCTA ATCTTCTTTC TCAGATTTGA
```

SEQ ID 12

EGLAIRMVALSLGCIIQCFDWQRLGEGLVDKTEGTGLTLPKAQPLVAKCSPRPIMANLLSQI

FIG. 7

SEQ ID 13 D56-AG10

```
   1 ATAGGTTTT GCGACTTTAG TGACACATCT GACTTTTGGT
  61 CGCTTGCTTC AAGGTTTTGA TTTTAGTAAG CCATCAAACA CGCCAATTGA CATGACAGAA
 121 GGCGTAGGCG TTACTTTGCC TAAGGTTAAT CAAGTTGAAG TTCTAATTAC CCCTCGTTTA
 181 CCTTCTAAGC TTTATTTATT TTGA
```

SEQ ID 14

IGFATLVTHLTFGRLLQGFDFSKPSNTPIDMTEGVGVTLPKVNQVEVLITPRLPSKLYLF

FIG. 8

SEQ ID 15 D35-33

```
   1 ATAGGCTTT GCGACTTTAG TGACACATCT GACTTTTGGT
  61 CGCTTGCTTC AAGGTTTTGA TTTTAGTAAG CCATCAAACA CGCCAATTGA CATGACAGAA
 121 GGCGTAGGCG TTACTTTGCC TAAGGTTAAT CAAGTTGAAG TTCTAATTAC CCCTCGTTTA
 181 CCTTCTAAGC TTTATTTAT
```

SEQ ID 16

IGFATLVTHLTFGRLLQGFDFSKPSNTPIDMTEGVGVTLPKVNQVEVLITPRLPSKLYL

FIG. 9

SEQ ID 17 D34-62

```
   1 ATAAATTTT GCGACTTTAG TGACACATCT GACTTTTGGT
  61 CGCTTGCTTC AAGGTTTTGA TTTTAGTACG CCATCAAACA CGCCAATAGA CATGACAGAA
 121 GGCGTAGGCG TTACTTTGCC TAAGGTAAAT CAAGTGGAAG TTCTAATTAG CCCTCGTTTA
 181 CCTTCTAAGC TTTATGTATT CTGA
```

SEQ ID 18

INFATLVTHLTFGRLLQGFDFSTPSNTPIDMTEGVGVTLPKVNQVEVLISPRLPSKLYVF

FIG. 10

SEQ ID 19 D56AA7

```
   1 ATTATACTT GCATTGCCAA TTCTTGGCAT CACTTTGGGA
  61 CGTTTGGTTC AGAACTTTGA GCTGTTGCCT CCTCCAGGCC AGTCGAAGCT CGACACCACA
 121 GAGAAAGGTG GACAGTTCAG TCTCCACATT TTGAAGCATT CCACCATTGT GTTGAAACCA
 181 AGGTCTTTCT GA
```

SEQ ID 20

IILALPILGITLGRLVQNFELLPPPGQSKLDTTEKGGQFSLHILKHSTIVLKPRSF

FIG. 11

SEQ ID 21 D56-AE1

```
   1 ATTATACTT GCATTGCCAA TTCTTGGCAT TACTTTGGGA
  61 CGTTTGGTTC AGAACTTTGA GCTGTTGCCT CCTCCAGGCC AGTCGAAGCT CGACACCACA
 121 GAGAAAGGTG GACAGTTCAG TCTCCATATT TTGAAGCATT CCACCATTGT GTTGAAACCA
 181 AGGTCTTGCT GA
```

SEQ ID 22

IILALPILGITLGRLVQNFELLPPPGQSKLDTTEKGGQFSLHILKHSTIVLKPRSC

FIG. 12

SEQ ID 23 D35-BB7

```
   1 TATTGCACTT GGGGTTGCAT CAATGGAACT TGCATTGTCA
  61 AATCTTCTTT ATGCATTTGA TTGGGAGTTA CCTTTTGGAA TGAAAAAAGA AGACATTGAC
 121 ACAAACGCCA GGCCTGGAAT TACCATGCAT AAGAAAAACG AACTTTATCT TATCCCTAAA
 181 AATTATCTAT AG
```

SEQ ID 24

IALGVASMELALSNLLYAFDWELPFGMKKEDIDTNARPGITMHKKNELYLIPKNYLPSKLYLF

FIG. 13

SEQ ID 25 D177-BA7

```
   1 ATTGCACTTG GGGTTGCATC CATGGAACTT
 121 GCTTTGTCAA ATCTTCTTTA TGCATTTGAT TGGGAGTTAC CTTACGGAGT GAAAAAAGAA
 181 AACATTGACA CAAATGTCAG GCCTGGAATT ACCATGCATA AGAAAAACGA ACTTTGCCTT
 241 ATCCCTAGAA ATTATCTATA G
```

SEQ ID 26

IALGVASMELALSNLLYAFDWELPYGVKKENIDTNVRPGITMHKKNELCLIPRNYL

FIG. 14

SEQ ID 27 D56A-AB6

```
   1 GGTATTGCAC TTGGGGTTGC ATCCATGGAA CTTGCTTTGT CAAATCTTCT TTATGCATTT
  61 GATTGGGAGT TGCCTTATGG AGTGAAAAAA GAAGACATCG ACACAAACGT TAGGCCTGGA
 121 ATTGCCATGC ACAAGAAAAA CGAACTTTGC CTTGTCCCAA AAAATTATTT ATAA
```

SEQ ID 28

IALGVASMELALSNLLYAFDWELPYGVKKEDIDTNVRPGIAMHKKNELCLVPKNYL

FIG. 15

SEQ ID 29 D144-AE2

```
   1 ATT GCACTTGGGG TTGCATCCAT GGAACTTGCT
  61 TTGTCAAATC TTCTTTATGC ATTTGATTGG GAGTTGCCTT ATGGAGTGAA AAAAGAAGAC
 121 ATCGACACAA ACGTTAGGCC TGGAATTGCC ATGCACAAGA AAAACGAACT TTGCCTTGTC
 181 CCAAAAAAAT TATTTATAAA TTATATTGGG ACGTGGATCT CATGCTAG
```

SEQ ID 30

IALGVASMELALSNLLYAFDWELPYGVKKEDIDTNVRPGIAMHKKNELCLVPKKLFINYIGTWISC

FIG. 16

SEQ ID 31 D56-AG11

```
   1 ATTTCGTTT GGTTTAGCTA ATGCTTATTT GCCATTGGCT
  61 CAATTACTTT ATCACTTTGA TTGGGAACTC CCCACTGGAA TCAAACCAAG CGACTTGGAC
 121 TTGACTGAGT TGGTTGGAGT AACTGCCGCT AGAAAAAGTG ACCTTTACTT GGTTGCGACT
 181 CCTTATCAAC CTCCTCAAAA CTGA
```

SEQ ID 32

ISFGLANAYLPLAQLLYHFDWELPTGIKPSDLDLTELVGVTAARKSDLYLVATPYQPPQN

FIG. 17

SEQ ID 33 D179-AA1

```
   1 ATTTCGTTT GGCTTAGCTA ATGCTTATTT GCCATTGGCT
  61 CAATTACTAT ATCACTTCGA TTGGAAACTC CCTGCTGGAA TCGAACCAAG CGACTTGGAC
 121 TTGACTGAGT TGGTTGGAGT AACTGCCGCT AGAAAAAGTG ACCTTTACTT GGTTGCGACT
 181 CCTTATCAAC CTCCTCAAAA GTGA
```

SEQ ID 34

ISFGLANAYLPLAQLLYHFDWKLPAGIEPSDLDLTELVGVTAARKSDLYLVATPYQPPQK

FIG. 18

SEQ ID 35 D56-AC7

```
   1 ATGCTATTT GGTTTAGCTA ATGTTGGACA ACCTTTAGCT
  61 CAGTTACTTT ATCACTTCGA TTGGAAACTC CCTAATGGAC AAAGTCATGA GAATTTCGAC
 121 ATGACTGAGT CACCTGGAAT TTCTGCTACA AGAAAGGATG ATCTTGTTTT GATTGCCACT
 181 CCTTATGATT CTTATTAATTCCAGTCTA TATCATCTAT ATGTACTCAA TAATTGTATG
 361 GGA
```

SEQ ID 36

MLFGLANVGQPLAQLLYHFDWKLPNGQSHENFDMTESPGISATRKDDLVLIATPYDSY

FIG. 19

SEQ ID 37 D144-AD1

```
   1 ATGC TATTTGGTTT AGCTAATGTT
  61 GGACAACCTT TAGCTCAGTT ACTTTATCAC TTCGATTGGA AACTCCCTAA TGGACAAACT
 121 CACCAAAATT TCGACATGAC TGAGTCACCT GGAATTTCTG CTACAAGAAA GGATGATCTT
 181 ATTTTGATTG CCACTCCTGC TCATTCTTGA
```

SEQ ID 38

MLFGLANVGQPLAQLLYHFDWKLPNGQTHQNFDMTESPGISATRKDDLILIATPAHS

FIG. 20

SEQ ID 39 D144-AB5

```
   1 TTAT TATTCGGTTT AGTTAATGTA
  61 GGACATCCTT TAGCTCAATT GCTTTATCAC TTCGATTGGA AGACTCTTCC TGGGATAAGT
 121 TCAGATAGTT TCGACATGAC TGAAACAGAT GGAGTAACTG CCGGAAGAAA GGATGATCTT
 181 TGTTTAATTG CTACTCCTTT TGGTCTCAAT TAA
```

SEQ ID 40

LLFGLVNVGHPLAQLLYHFDWKTLPGISSDSFDMTETDGVTAGRKDDLCLIATPFGLN

FIG. 21

SEQ ID 41 D181-AB5

```
  1 A TGTCGTTTGG TTTAGTTAAC ACTGGGCATC CTTTAGCTCA
 61 GTTGCTCTAT TTCTTTGACT GGAAATTCCC TCATAAGGTT AATGCAGCTG ATTTTCACAC
121 TACTGAAACA AGTAGAGTTT TTGCAGCAAG CAAAGATGAC CTCTACTTGA TTCCAACAAA
181 TCACATGGAG CAAGAGTAG
```

SEQ ID 42

MSFGLVNTGHPLAQLLYFFDWKFPHKVNAADFHTTETSRVFAASKDDLYLIPTNHMEQE

FIG. 22

SEQ ID 43 D73-AC9

```
  1 AT GTCGTTTGGT TTAGTTAACA CAGGGCATCC TTTAGCCCAG
121 TTGCTCTATT GCTTTGACTG GAAACTCCCT GACAAGGTTA ATGCAAATGA TTTTCGCACT
181 ACTGAAACAA GTAGAGTTTT TGCAGCAAGC AAAGATGACC TCTACTTGAT TCCCACAAAT
241 CACAGGGAGC AAGAATAG
```

SEQ ID 44

MSFGLVNTGHPLAQLLYCFDWKLPDKVNANDFRTTETSRVFAASKDDLYLIPTNHREQE

FIG. 23

SEQ ID 45 D56-AC12

```
  1 ATGCAATTT GGTTTGGCTC TTGTTACTCT GCCATTGGCT
 61 CATTTGCTTC ACAATTTTGA TTGGAAACTT CCCGAAGGAA TTAATGCAAG GGATTTGGAC
121 ATGACAGAGG CAAATGGGAT ATCTGCTAGA AGAGAAAAAG ATCTTTACTT GATTGCTACT
181 CCTTATGTAT CACCTCTTGA TTAA
```

SEQ ID 46

MQFGLALVTLPLAHLLHNFDWKLPEGINARDLDMTEANGISARREKDLYLIATPYVSPLD

FIG. 24

SEQ ID 47 D58-AB9

```
  1 ATGACTTAT GCATTGCAAG TGGAACACCT AACAATGGCA
 61 CATTTGATCC AGGGTTTCAA TTACAGAACT CCAACTGATG AGCCCTTGGA TATGAAAGAA
121 GGTGCAGGCA TAACTATACG TAAGGTAAAT CCTGTGAAAG TGATAATTAC GCCTCGCTTG
181 GCACCTGAGC TTTATTAA
```

SEQ ID 48

MTYALQVEHLTMAHLIQGFNYRTPTDEPLDMKEGAGITIRKVNPVKVIITPRLAPELY

FIG. 25

SEQ ID 49 D56-AG9

```
  1 ATGACTTAT GCATTGCAAG TGGAACACCT AACAATGGCA
 61 CATTTAATCC AGGGTTTCAA TTACAAAACT CCAAATGACG AGGCCTTGGA TATGAAGGAA
121 GGTGCAGGCA TAACTATACG TAAGGTAAAT CCTGTGGAAC TGATAATAGC GCCTCGCCTG
181 GCACCTGAGC TTTATTAA
```

SEQ ID 50

MTYALQVEHLTMAHLIQGFNYKTPNDEALDMKEGAGITIRKVNPVELIIAPRLAPELY

FIG. 26

SEQ ID 51 D56-AG6

```
   1 ATGACTTAT GCATTGCAAG TGGAACACCT AACAATGGCA
  61 CATTTAATCC AGGGTTTCAA TTACAAAACT CCAAATGACG AGGCCTTGGA TATGAAGGAA
 121 GGTGCAGGCA TAACAATACG TAAGGTAAAT CCAGTGGAAT TGATAATAAC GCCTCGCTTG
 181 GCACCTGAGC TTTACTAA
```

SEQ ID 52

MTYALQVEHLTMAHLIQGFNYKTPNDEALDMKEGAGITIRKVNPVELIITPRLAPELY

FIG. 27

SEQ ID 53 D35-BG11

```
   1 ATGACTTAT GCATTGCAAG TGGAACACTT AACAATGGCA
  61 CATTTGATCC AAGGTTTCAA TTACAGAACT CCAAATGACG AGCCCTTGGA TATGAAGGAA
 121 GGTGCAGGCA TAACTATACG TAAGGTAAAT CCTGTGGAAC TGATAATAGC GCCTCGCCTG
 181 GCACCTGAGC TTTATTAA
```

SEQ ID 54

MTYALQVEHLTMAHLIQGFNYRTPNDEPLDMKEGAGITIRKVNPVELIIAPRLAPELY

FIG. 28

SEQ ID 55 D35-42

```
   1 ATGACTTAT GCATTGCAAG TGGAACACTT AACAATGGCA
  61 CATTTGATCC AAGGTTTCAA TTACAGAACT CCAAATGACG AGCCCTTGGA TATGAAGGAA
 121 GGTGCAGGCA TAACTATACG TAAGGTAAAT CCTGTGGAAC TGATAATAGC GCCCCTGGCA
 181 CCTGAGCTTT ATTAA
```

SEQ ID 56

MTYALQVEHLTMAHLIQGFNYRTPNDEPLDMKEGAGITIRKVNPVELIIAPLAPELY

FIG. 29

SEQ ID 57 D35-BA3

```
   1 ATGACTTAT GCATTGCAAG TGGAACACTT AACAATGGCA
  61 CATTTGATCC AAGGTTTCAA TTACAGAACT CCAAATGACG AGCCCTTGGA TATGAAGGAA
 121 GGTGCAGGCA TAACTATACG TAAGGTAAAT CCTGCGGAAC TGATAATAGC GCCTCGCCTG
 181 GCACCTGAGC TTTATTAA
```

SEQ ID 58

MTYALQVEHLTMAHLIQGFNYRTPNDEPLDMKEGAGITIRKVNPAELIIAPRLAPELY

FIG. 30

SEQ ID 59 D34-57

```
   1 ATGACTTAT GCATTACAAG TGGAACACCT AACAATAGCA
  61 CATTTGATCC AGGGTTTCAA TTACAAAACT CCAAATGACG AGCCCTTGGA TATGAAGGAA
 121 GGTGCAGGAT TAACCATACG TAAAGTAAAT CCTGTAGAAG TGACAACTAC GGCTCGCCTG
 181 GCACCTGAGC TTTATTAA
```

SEQ ID 60

MTYALQVEHLTIAHLIQGFNYKTPNDEPLDMKEGAGLTIRKVNPVEVTTTARLAPELY

FIG. 31

SEQ ID 61 D34-52

```
   1 ATGACTTAT GCATTACAAG TGGAACACCT AACAATAGCA
  61 CATTTGATCC AGGGTTTCAA TTACAAAACT CCAAATGACG AGCCCTTGGA TATGAAGGAA
 121 GGTGCAGGAT TAACTATACG TAAAGTAAAT CCTGTAGAAG TGACAATTAC GGCTCGCCTG
 181 GCACCTGAGC TTTATTAA
```

SEQ ID 62

MTYALQVEHLTIAHLIQGFNYKTPNDEPLDMKEGAGLTIRKVNPVEVTITARLAPELY

FIG. 32

SEQ ID 63 D34-25

```
   1 ATGACTTAT GCATTACAAG TGGAACACCT AACAATAGCA
  61 CATTTGATCC AGGGTTTCAA TTACAAAACT CCAAATGACG AGCCCCTGGA TATGAAGGAA
 121 GGTGCAGGAT TAACTATACG TAAAGTAAAT CCTGTAGAAG TGACAATTAC GGCTCGCCTG
 181 GCACCTGAGC TTTATTAA
```

SEQ ID 64

MTYALQVEHLTIAHLIQGFNYKTPNDEPLDMKEGAGLTIRKVNPVEVTITARLAPELY

FIG. 33

SEQ ID 65 D56AD10

```
   1 TATAGCCTT GGACTTAAGG TTATCCGAGT AACATTAGCC
  61 AACATGTTGC ATGGATTCAA CTGGAAATTA CCTGAAGGTA TGAAGCCAGA AGATATAAGT
 121 GTGGAAGAAC ATTATGGGCT CACTACACAT CCTAAGTTTC CTGTTCCTGT GATCTTGGAA
 181 TCTAGACTTT CTTCAGATCT CTATTCCCCC ATCACTTAA
```

SEQ ID 66

YSLGLKVIRVTLANMLHGFNWKLPEGMKPEDISVEEHYGLTTHPKFPVPVILESRLSSDLYSPIT

FIG. 34

SEQ ID 67 D56-AA11

```
   1 ATACAGTCTT GGGATTCGTA TAATTAGGGC AACTTTAGCT
  61 AACTTGTTGC ATGGATTCAA CTGGAGATTG CCTAATGGTA TGAGTCCAGA AGACATTAGC
 121 ATGGAAGAGA TTTATGGGCT AATTACACAC CCCAAAGTCG CACTTGACGT GATGATGGAG
 181 CCTCGACTTC CCAACCATCT TTACAAATAG
```

SEQ ID 68

YSLGIRIIRATLANLLHGFNWRLPNGMSPEDISMEEIYGLITHPKVALDVMMEPRLPNHLYK

FIG. 35

SEQ ID 69 D177-BD5

```
   1 ATTAATTTTT CAATACCACT TGTTGAGCTT
 121 GCACTTGCTA ATCTATTGTT TCATTATAAT TGGTCACTTC CTGAAGGGAT GCTAGCTAAG
 181 GATGTTGATA TGGAAGAAGC TTTGGGGATT ACCATGCACA AGAAATCTCC CCTTTGCTTA
 241 GTAGCTTCTC ATTATACTTG TTGA
```

SEQ ID 70

INFSIPLVELALANLLFHYNWSLPEGMLAKDVDMEEALGITMHKKSPLCLVASHYTC

FIG. 36

SEQ ID 71 D56A-AG10

```
   1 ATGCAACTTG GGCTTTATGC ATTGGAAATG GCTGTGGCCC ATCTTCTTCA TTGTTTTACT
  61 TGGGAATTGC CAGATGGTAT GAAACCAAGT GAGCTTAAAA TGGATGATAT TTTTGGACTC
 121 ACTGCTCCAA AAGCTAATCG ACTCGTGGCT GTGCCTACTC CACGTTTGTT GTGTCCCCTT
 181 TATTAATTGA
```

SEQ ID 72

MQLGLYALEMAVAHLLHCFTWELPDGMKPSELKMDDIFGLTAPKANRLVAVPTPRLLCPLY

FIG. 37

SEQ ID 73 58-BC5

```
   1 ATGCAACTT GGGCTTTATG CATTAGAAAT GGCAGTGGCC
  61 CATCTTCTTC TTTGCTTTAC TTGGGAATTG CCAGATGGTA TGAAACCAAG TGAGCTTAAA
 121 ATGGATGATA TTTTTGGACT CACTGCTCCA AGAGCTAATC GACTCGTGGC TGTGCCTAGT
 181 CCACGTTTGT TGTGCCCACT TTATTAA
```

SEQ ID 74

MQLGLYALEMAVAHLLLCFTWELPDGMKPSELKMDDIFGLTAPRANRLVAVPSPRLLCPLY

FIG. 38

SEQ ID 75 D58-AD12

```
   1 ATGCAACTT GGGCTTTATG CATTGGAAAT GGCTGTGGCC
  61 CATCTTCTTC ATTGTTTTAC TTGGGAATTG CCAGATGGTA TGAAACCAAG TGAGCTTAAA
 121 ATGGATGATA TTTTTGGACT CACTGCTCCA AGAGCTAATC GACTCGTGGC TGTGCCTACT
 181 CCACGTTTGT TGTGTCCCCT TTATTAA
```

SEQ ID 76

MQLGLYALEMAVAHLLHCFTWELPDGMKPSELKMDDIFGLTAPRANRLVAVPTPRLLCPLY

FIG. 39

SEQ ID 77 D56-AC11

```
   1 ATGCTTTGG AGTGCGAGTA TAGTGCGCGT CAGCTACCTA
  61 ACTTGTATTT ATAGATTCCA AGTATATGCT GGGTCTGTGT TCAGAGTAGC ATGA
```

SEQ ID 78

MLWSASIVRVSYLTCIYRFQVYAGSVFRVA

FIG. 40

SEQ ID 79 D35-39

```
   1 ATGCTTTGG AGTGCGAGTA TAGTGCGCGT CAGCTACCTA
  61 ACTTGTATTT ATAGATTCCA AGTATATGCT GGGTCTGTGT TCAGAGTAGC ATGA
```

SEQ ID 80

MLWSASIVRVSYLTCIYRFQVYAGSVFRVA

FIG. 41

SEQ ID 81 D58-BH4

```
   1 ATGCTTTGG AGTGCGAGTA TAGTGCGCGT CAGCTACCTA
  61 ACCTGTATTT ATAGATTCCA AGTATATGCT GGGTCTGTGT TCAGAGTAGC ATGA
```

SEQ ID 82
MLWSASIVRVSYLTCIYRFQVYAGSVFRVA

FIG. 42

SEQ ID 83 D177-BD7

```
   1 ATTAATTTTT CAATACCACT TGTTGAGCTT GCACTTGCTA ATCTATTGTT TCATTATAAT
  61 TGGTCACTTC CTGAGGGGAT GCTACCTAAG GATGTTGATA TGGAAGAAGC TTTGGGGATT
 121 ACCATGCACA AGAAATCTCC CCTTTGCTTA GTAGCTTCTC ATTATAACTT GTTGTGA
```

SEQ ID 84
INFSIPLVELALANLLFHYNWSLPEGMLPKDVDMEEALGITMHKKSPLCLVASHYNLL

FIG. 43

SEQ ID 85
D176-BF2

```
   1 AT ATCATTTGGT TTGGCTAATG TTTATTTGCC ACTAGCTCAA
 121 TTGTTATATC ATTTTGATTG GAAACTCCCT ACTGGAATCA ATTCAAGTGA CTTGGACATG
 181 ACTGAGTCGT CAGGAGTAAC TTGTGCTAGA AAGAGTGATT TATACTTGAC TGCTACTCCA
 241 TATCAACTTT CTCAAGAGTG A
```

SEQ ID 86
GISFGLANVYLPLAQLLYHFDWKLPTGINSSDLDMTESSGVTCARKSDLYLTATPYQLSQE

FIG. 44

SEQ ID 87 D56-AD6

```
   1 ATGCTTTGG AGTGCGAGTA TAGTGCGCGT CAGCTACCTA
  61 ACTTGTATTT ATAGATTCCA AGTATATGCT GGGTCTGTGT CCAGAGTAGC ATGA
```

SEQ ID 88
MLWSASIVRVSYLTCIYRFQVYAGSVSRVA

FIG. 45

SEQ ID 89 D73A-AD6

```
   1 CT GAATTTTGCA ATGTTAGAGG CAAAAATGGC ACTTGCATTG
 121 ATTCTACAAC ACTATGCTTT TGAGCTCTCT CCATCTTATG CACATGCTCC TCATACAATT
 181 ATCACTCTGC AACCTCAACA TGGTGCTCCT TTGATTTTGC GCAAGCTGTA G
```

SEQ ID 90
LNFAMLEAKMALALILQHYAFELSPSYAHAPHTIITLQPQHGAPLILRKL

FIG. 46

SEQ ID 91 D70A-BA11

```
   1 CT GAATTTTGCA ATGTTAGAGG CAAAAATGGC ACTTGCATTG
 121 ATTCTACAAC ACTATGCTTT TGAGCTCTCT CCATCTTATG CACACGCTCC TCATACAATT
 181 ATCACTCTGC AACCTCAACA TGGTGCTCCT TTGATTTTGC GCAAGCTGTA G
```

SEQ ID 92

LNFAMLEAKMALALILQHYAFELSPSYAHAPHTIITLQPQHGAPLILRKL

FIG. 47

SEQ ID 93 D70A-BB5

```
   1 AA TAATTTTGCA ATGTTGGAAA CTAAGATTGC CTTAGCAATG
 121 ATCCTACAGC GTTTTGCTTT CGAGCTTTCT CCATCTTACG CTCATGCACC TACTTATGTC
 181 GTCACTCTTC GACCTCAGTG TGGTGCTCAC TTAATCTTGC AAAAATTATA GGTCCTTAAT
 241 CTGGATTTCC CATTATTGAG TAGTGCCTAA TAAATCTTCT CTATCACTAT TTTTCCATCT
 301 TTCA
```

SEQ ID 94

NNFAMLETKIALAMILQRFAFELSPSYAHAPTYVVTLRPQCGAHLILQKL

FIG. 48

SEQ ID 95 D70A-AB5

```
   1 AGCGAAGGGG TGGCAAAGGC AACAAAGGGG AAAATGACAT ATTTTCCATT TGGTGCAGGA
  61 CCGCGAAAAT GCATTGGGCA AAACTTCGCG ATTTTGGAAG CAAAAATGGC TATAGCTATG
 121 ATTCTACAAC GCTTCTCCTT CGAGCTCTCC CCATCTTATA CACACTCTCC ATACACTGTG
 181 GTCACTTTGA AACCCAAATA TGGTGCTCCC CTAATAATGC ACAGGCTGTA GTCCTGTGAG
 241 AATATGCTAT CCGAGGAATT CAGTTCCT
```

SEQ ID 96

QNFAILEAKMAIAMILQRFSFELSPSYTHSPYTVVTLKPKYGAPLIMHRL

FIG. 49

SEQ ID 97 D70A-AA8

```
   1 AGCGAAGGGG TGGCAAAGGC AACAAAGGGG AAAATGACAT ATTTTCCATT TGGTGCAGGA
  61 CCGCGAAAAT GCATTGGGCA AAACTTCGCG ATTTTGGAAG CAAAAATGGC TATAGCTATG
 121 ATTCTACAAC GCTTCTCCTT CGAGCTCTCT CCATCTTATA CACACTCTCC ATACACTGTG
 181 GTCACTTTGA AACCCAAATA TGGTGCTCCC CTAATAATGC ACAGGCTGTA GTCCTGT
```

SEQ ID 98

QNFAILEAKMAIAMILQRFSFELSPSYTHSPYTVVTLKPKYGAPLIMHRL

FIG. 50

SEQ ID 99 D70A-AB8

```
   1 C AAAATTTTGC CATGTTAGAA GCAAAGATGG CTCTGTCTAT GATCCTGCAA
 121 CGCTTCTCTT TTGAACTGTC TCCGTCTTAT GCACATGCCC CTCAGTCCAT ATTAACCGT
 181 CAGCCACAAT ATGGTGCTCC ACTTATTTTC CACAAGCTAT AA
```

SEQ ID 100

QNFAMLEAKMALSMILQRFSFELSPSYAHAPQSILTVQPQYGAPLIFHKL

FIG. 51

SEQ ID 101 **D70A-BH2**

```
  1 AT AAACTTTGCA ATGACAGAAG CGAAGATGGC TATGGCTATG
121 ATTCTGCAAC GCTTCTCCTT TGAGCTATCT CCATCTTACA CACATGCTCC ACAGTCTGTA
181 ATAACTATGC AACCCCAATA TGGTGCTCCT CTTATATTGC ACAAATTGTA A
```

SEQ ID 102

INFAMTEAKMAMAMILQRFSFELSPSYTHAPQSVITMQPQYGAPLILHKL

FIG. 52

SEQ ID 103 **D70A-AA4**

```
  1 AT AAACTTTGCA ATGGCAGAAG CGAAGATGGC TATGGCTATG
121 ATTCTGCAAC GCTTCTCCTT TGAGCTATCT CCATCTTACA CACATGCTCC ACAGTCTGTA
181 ATAACTATGC AACCCCAATA TGGTGCTCCT CTTATATTGC ACAAATTGTA A
```

SEQ ID 104

INFAMAEAKMAMAMILQRFSFELSPSYTHAPQSVITMQPQYGAPLILHKL

FIG. 53

SEQ ID 105 **D70A-BA1**

```
  1 CA AAACTTTGCA ATGATGGAAG CAAAAATGGC AGTAGCTATG
121 ATACTACAAA AATTTTCCTT TGAACTATCC CCTTCTTATA CACATGCTCC ATTTGCAATT
181 GTGACTATTC ATCCTCAGTA TGGTGCTCCT CTGCTTATGC GCAGACTTTA A
```

SEQ ID 106

QNFAMMEAKMAVAMILQKFSFELSPSYTHAPFAIVTIHPQYGAPLLMRRL

FIG. 54

SEQ ID 107 **D70A-BA9**

```
  1 CA AAACTTTGCA ATGATGGAAG CAAAAATGGC AGTAGCTATG
121 ATACTACATA AATTTTCCTT TGAACTATCC CCTTCTTATA CACATGCTCC ATTTGCAATT
181 GTGACTATTC ATCCTCAGTA TGGTGCTCCT CTGCTTATGC GCAGACTTTA A
```

SEQ ID 108

QNFAMMEAKMAVAMILHKFSFELPSYTHAPFAIVTIHPQYGAPLLMRRL

FIG. 55

SEQ ID 109 **D70A-BD4**

```
  1 CA AAATTTTGCT ATGTTAGAGG CTAAAATGGC AATGGCTATG
121 ATTCTGAAAA CCTATGCATT TGAACTCTCT CCATCTTATG CTCATGCTCC TCATCCACTA
181 CTACTTCAAC CTCAATATGG TGCTCAATTA ATTTTGTACA AGTTGTAG
```

SEQ ID 110

QNFAMLEAKMAMAMILKTYAFELSPSYAHAPHPLLLQPQYGAQLILYKL

FIG. 56

SEQ ID 111 **D181-AC5**

```
   1 TATAGCATGG GGCTCAAGGC GATTCAAGCT AGCTTAGCTA
  61 ATCTTCTACA TGGATTTAAC TGGTCATTGC CTGATAATAT GACTCCTGAG GACCTCAACA
 121 TGGATGAGAT TTTTGGGCTC TCTACACCTA AAAAATTTCC ACTTGCTACT GTGATTGAGC
 181 CAAGACTTTC ACCAAAACTT TACTCTGTTT GA
```

SEQ ID 112

YSMGLKAIQASLANLLHGFNWSLPDNMTPEDLNMDEIFGLSTPKKFPLATVIEPRLSPKLYSV

FIG. 57

SEQ ID 113 **D144-AH1**

```
   1 TAT AGCTTGGGGC TCAAGGAGAT TCAAGCTAGC
  61 TTAGCTAATC TTCTACATGG ATTTAACTGG TCATTGCCTG ATAATATGAC TCCTGAGGAC
 121 CTCAACATGG ATGAGATTTT TGGGCTCTCT ACACCTAAAA AATTTCCACT TGCTACTGTG
 181 ATTGAGCCAA GACTTTCACC AAAACTTTAC TCTGTTTGA
```

SEQ ID 114

YSLGLKEIQASLANLLHGFNWSLPDNMTPEDLNMDEIFGLSTPKKFPLATVIEPRLSPKLYSV

FIG. 58

SEQ ID 115 **D34-65**

```
   1 CATAGCTTG GGGCTCAAGG TGATTCAAGC TAGCTTAGCT
  61 AATCTTCTAC ATGGATTTAA CTGGTCATTG CCTGATAATA TGACTCCTGA GGACCTCAAC
 121 ATGGATGAGA TTTTTGGGCT CTCTACACCT AAAAAATTTC CACTTGCTAC TGTGATTGAG
 181 CCAAGACTTT CACCAAAACT TTACTCTGTT TGA
```

SEQ ID 116

HSLGLKVIQASLANLLHGFNWSLPDNMTPEDLNMDEIFGLSTPKKFPLATVIEPRLSPKLYSV

FIG. 59

SEQ ID 117 **D35-BG2**

```
   1 CTGTGCTTT CCATGTTTAA TCTCTAGTTA TATACTGGCT
  61 TTGAATGTGA ATCTGTATCA TAATTTCTTG CAAATTTCTC CTTCCATTTC TTATTAA
```

SEQ ID 118

LCFPCLISSYILALNVNLYHNFLQISPSISY

FIG. 60

SEQ ID 119 **D73A-AH7**

```
   1 TCTG GACTTGCTCA ATGTGTGGTT GGTTTAGCTT TAGCAACTCT AGTGCAGTGT
 121 TTTGAGTGGA AAAGGGTAAG CGAAGAGGTG GTTGATTTGA CGGAAGGAAA AGGTCTCACT
 181 ATGCCAAAAC CCGAGCCACT CATGGCTAGG TGCGAAGCTC GTGACATTTT TCACAAAGTT
 241 CTTTCAGAAA TATCTTAA
```

SEQ ID 120

SGLAQCVVGLALATLVQCFEWKRVSEEVVDLTEGKGLTMPKPEPLMARCEARDIFHKVLSEIS

FIG. 61

SEQ ID 121 D58-AA1

```
   1 TTGGGCTTG GCAACGGTGC ATGTGAATTT GATGTTGGCC
  61 CGAATGATTC AAGAATTTGA ATGGTCCGCT TACCCGGAAA ATAGGAAAGT GGATTTTACT
 121 GAGAAATTGG AATTTACTGT GGTGATGAAA AATCCTTTAA GAGCTAAGGT CAAGCCAAGA
 181 ATGCAAGTGG TGTAA
```

SEQ ID 122

LGLATVHVNLMLARMIQEFEWSAYPENRKVDFTEKLEFTVVMKNPLRAKVKPRMQVV

FIG. 62

SEQ ID 123 D73A-AE10

```
   1 TATGCTT TGGCTATGCT TCATTTAGAG
 121 TACTTTGTGG CTAATTTGGT TTGGCATTTT CGATGGGAGG CTGTGGAGGG AGATGATGTT
 181 GATCTTTCAG AAAAGCTAGA ATTCACCGTT GTGATGAAGA ATCCACTTCG AGCTCGTATC
 241 TGCCCCAGAG TTAACTCTAT TTGA
```

SEQ ID 124

YALAMLHLEYFVANLVWHFRWEAVEGDDVDLSEKLEFTVVMKNPLRARICPRVNSI

FIG. 63

SEQ ID 125 D56A-AC12

```
   1 GGTCAGCAAG TTGGACTTCT TAGAACAACC ATTTTCATCG CCTCATTACT GTCTGAATAT
  61 AAGCTGAAAC CTCGCTCACA CCAGAAACAA GTTGAACTCA CCGATTTAAA TCCAGCAAGT
 121 TGGCTTCATT CGATAAAAGG CGAACTGTTA GTCGATGCGA TTCCTCGAAA GAAGGCGGCA
 181 TTTTAA
```

SEQ ID 126

GQQVGLLRTTIFIASLLSEYKLKPRSHQKQVELTDLNPASWLHSIKGELLVDAIPRKKAAF

FIG. 64

SEQ ID 127 D177-BF7

```
   1 ATCACATTTG CTAAGTTTGT GAATGAGCTA
 121 GCATTGGCAA GATTAATGTT CCATTTTGAT TTCTCGCTAC CAAAAGGAGT TAAGCATGAG
 181 GATTTGGACG TGGAGGAAGC TGCTGGAATT ACTGTTAGAA GGAAGTTCCC CCTTTTAGCC
 241 GTCGCCACTC CATGCTCGTG A
```

SEQ ID 128

ITFAKFVNELALARLMFHFDFSLPKGVKHEDLDVEEAAGITVRRKFPLLAVATPCS

FIG. 65

SEQ ID 129 D73A-AG3

```
   1 CA GAGGTATGCT ATAAACCATT TGATGCTCTT TATTGCGTTG
 121 TTCACGGCTC TGATTGATTT CAAGAGGCAC AAAACGGACG GCTGTGATGA TATCGCGTAT
 181 ATTCCAACCA TTGCTCCAAA GGATGATTGT AAAGTGTTCC TTTCACAGAG GTGCACTCGA
 241 TTCCCATCTT TTTCATGA
```

SEQ ID 130

QRYAINHLMLFIALFTALIDFKRHKTDGCDDIAYIPTIAPKDDCKVFLSQRCTRFPSFS

FIG. 66

SEQ ID 131 **D70A-AA12**

```
  1 ATG TCATTTGGTT TAGCTAATCT TTACTTACCA TTGGCTCAAT
121 TACTCTATCA CTTTGACTGG AAACTCCCAA CCGGAATCAA GCCAAGAGAC TTGGACTTGA
181 CCGAATTATC GGGAATAACT ATTGCTAGAA AGGGTGACCT TTACTTAAAT GCTACTCCTT
241 ATCAACCTTC TCGAGAGTAA
```

SEQ ID 132

MSFGLANLYLPLAQLLYHFDWKLPTGIKPRDLDLTELSGITIARKGDLYLNATPYQPSRE

FIG. 67

SEQ ID 133 **D185-BC1**

```
  1 TTGGGCTTG GCAACGGTGC ATGTGAATTT GATGTTGGCC
 61 CGAACGATTC AAGAATTTGA ATGGTCCGCT TACCCGGAAA ATAGGAAAGT GGATTTtACT
121 GAGAAATTGG AATTTACTGT GGTGATGAAA AACCCTTTAA GAGCTAAGGT CAAGCCAAGA
181 ATGCAAGTGG TGTAA
```

SEQ ID 134

LGLATVHVNLMLARTIQEFEWSAYPENRKVDFTEKLEFTVVMKNPLRAKVKPRMQVV

FIG. 68

SEQ ID 135 **D185-BG2**

```
  1 TTGGGCTTG GCAACGGTGC ATGTGAATTT GATGTTGGCC
 61 CGAATGATTC AAGAATTTGA ATGGTCCGCT TACCCGGAAA ATAGGAAAGT GGATTTACTG
121 AGAAATTGGA ATTTACTGTG GTGA
```

SEQ ID 136

LGLATVHVNLMLARMIQEFEWSAYPENRKVDLLRNWNLLW

FIG. 69

SEQ ID 137 **D185-BE1**

```
  1 ATCACATTT GCTAAGTTTG TGAATGAGCT AGCATTGGCA
 61 AGATTAATGT TCCATTTTGA TTTCTCGCTA CCAAAAGGAG TTAAGCATGA GGATTTGGAC
121 GTGGAGGAAG CTGCTGGAAT TACTGTTAGG AGGAAGTTCC CCCTTTTAGC CGTCGCCACT
181 CCATGCTCGT GA
```

SEQ ID 138

ITFAKFVNELALARLMFHFDFSLPKGVKHEDLDVEEAAGITVRRKFPLLAVATPCS

FIG. 70

SEQ ID 139 **D185-BD2**

```
  1 ATCACATTT GCTAAGTTTG TGAATGAGCT AGCATTGGCA
 61 AGATTAATGT TCCATTTTGA TTTCTCGCTA CCAAAAGGAG TTAAGCATGC GGATTTGGAC
121 GTGGAGGAAG CTGCTGGAAT TACTGTTAGA AGGAAGTTCC CCCTTTTAGC CGTCGCCACT
181 CCATGCTCGT GA
```

SEQ ID 140

ITFAKFVNELALARLMFHFDFSLPKGVKHADLDVEEAAGITVRRKFPLLAVATPCS

FIG. 71

SEQ ID 141 D176-BG2

```
  1 CA AAATTTTGCC ATGTTAGAAG CAAAGACTAC TTTGGCTATG
121 ATCCTACAAC GCTTCTCCTT TGAACTGTCT CCATCTTATG CACATGCTCC TCAGTCCATA
181 ATAACTTTGC AACCCCAGTA TGGTGCTCCA CTTATTTTGC ATAAAATATA G
```

SEQ ID 142

QNFAMLEAKTTLAMILQRFSFELSPSYAHAPQSIITLQPQYGAPLILHKI

FIG. 72

SEQ ID 143 D185-BD3

```
  1 ATTATCCTT GCACTGCCAA TTCTTGGCAT TACCTTGGGA
 61 CGCTTGGTGC AGAACTTTGA GTTGTTGCCT CCTCCAGGAC AGTCAAAGCT TGACACAACA
121 GAGAAAGGCG GGCAATTCAG TCTGCACATT TTGAAGCATT CCACCATTGT GATGAAACCA
181 AGATCTTTTT AA
```

SEQ ID 144

IILALPILGITLGRLVQNFELLPPPGQSKLDTTEKGGQFSLHILKHSTIVMKPRSF

FIG. 73

SEQ ID 145 D176-BC3

```
  1 C AAAATTTTGC CATGTTAGAA GCAAAGACTA CTTTGGCTAT
121 GATCCTACAA CGCTTCTCCT TTGAACTGTC TCCATCTTAT GCACATGCTC CTCAGTCCAT
181 AATAACTTGC AACCCCAGTA TGGTGCTCCA CTTATTTTGC ATAAAATATA GTTTATTACT
241 TGTAAGTAGT GTCTCGTTTT ATGTTAAGCA TGAGTCCAAA ATGTTAAGGC TTGTAGAACT
301 GCAAAATGGG AATGCATTTG CACTCGTGCA CTGTAGATTG TTGTAA
```

SEQ ID 146

QNFAMLEAKTTLAMILQRFSFELSPSYAHAPQSIITCNPSMVLHLFCIKYSLLLVSSVSFYVKHESKMLRLVELQNGNA
FALVHCRLL

FIG. 74

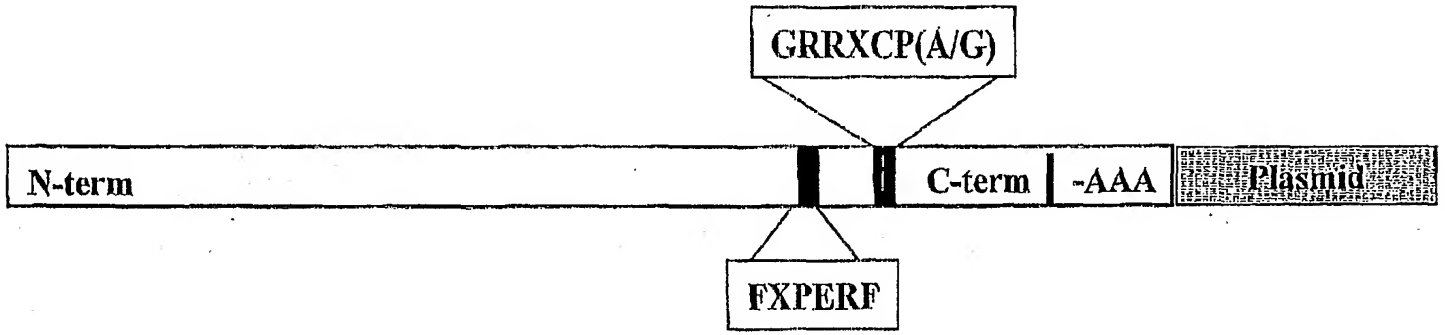
SEQ ID 147 D176-BB3

```
  1 GCTGAT
 61 ATGGGGTTGC GAGCAGTTTC TTTGGCATTA GGTGCACTTA TTCAATGCTT TGACTGGCAA
121 ATTGAGGAAG CGGAAAGCTT GGAGGAAAGC TATAATTCTA GAATGACTAT GCAGAACAAG
181 CCTTTGAAGG TTGTCTGCAC TCCACGCGAA GATCTTGGCC AGCTTCTATC CCAACTCTAA
```

SEQ ID 148

ADMGLRAVSLALGALIQCFDWQIEEAESLEESYNSRMTMQNKPLKVVCTPREDLGQLLSQL

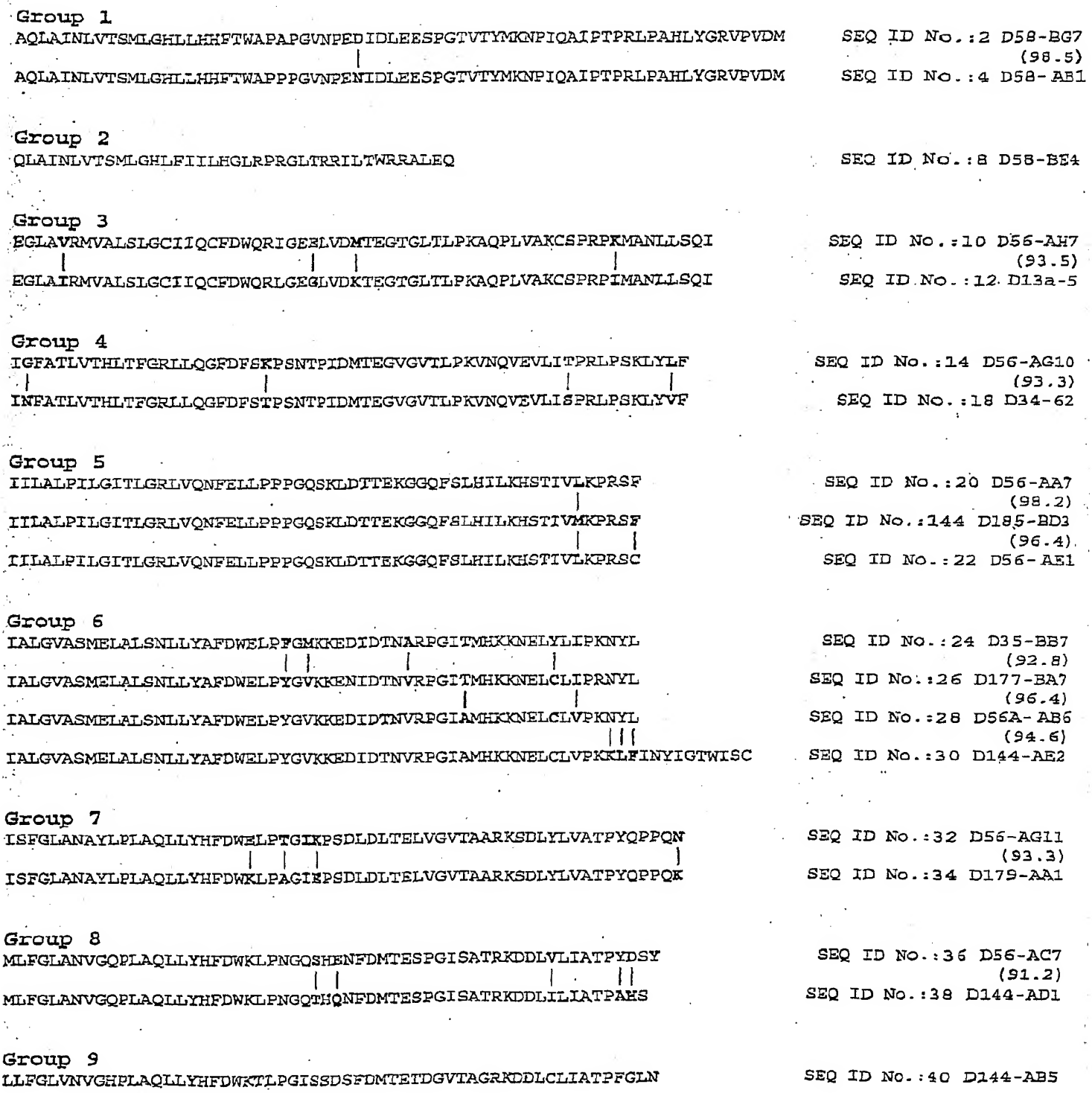
Figure 75 Cloning of Cytochrome P450 cDNA fragments by PCR



DM	FXPERF -for	5'-TTYIIICCIGARMGITTY-3'	SEQ ID NO. 149
DM4	GRRXCP(A/G)-for	5'-GGIMGIMGIIIITGYCCIGS-3'	SEQ ID NO. 150
DM12	FKPERF-for	5'-TTYAARCCTGAGAGATT-3'	SEQ ID NO. 151
DM13	PERFL-for	5'-CCAGARAGATTCTTG-3'	SEQ ID NO. 152
DM17	GRRMCP-for	5'-GGRMGRMGRATGTGYCC-3'	SEQ ID NO. 153
OLIGO d(T)		5'-TTTTTTTTTTTTTTTTTTN-3'	SEQ ID NO. 154
T7		5'-ATTATGCTGAGTGATATCCC-3'	SEQ ID NO. 155
SP6		5'-ATTTAGGTGACACTATAG-3'	SEQ ID NO. 156

I = Deoxyinosine; Y = C, T; M = A,C; R = A,G; S = C,G; N = A,T,C,G

Figure 74: Amino Acid Identity of Group Members



76

Figure : Amino Acid Identity of Group Members

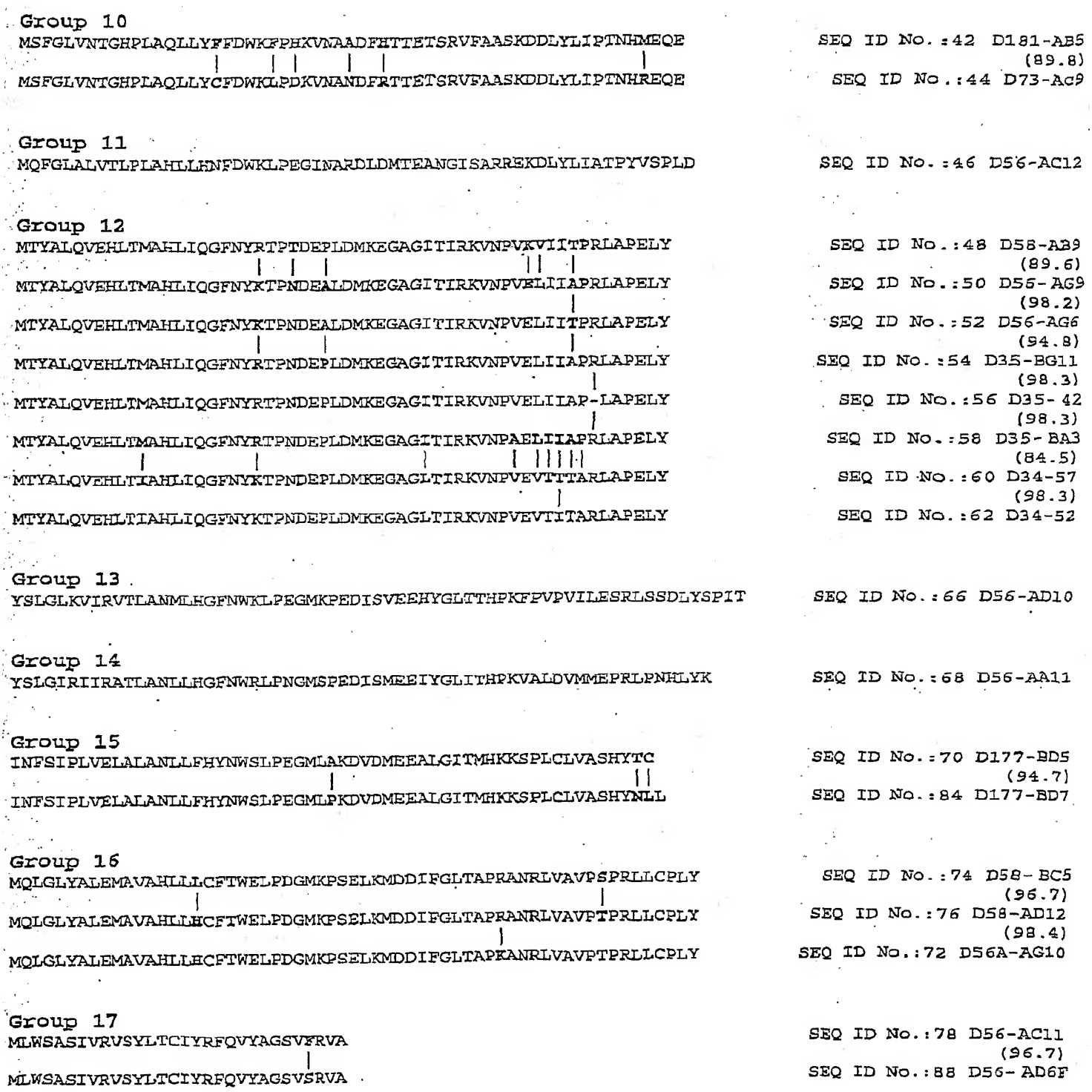
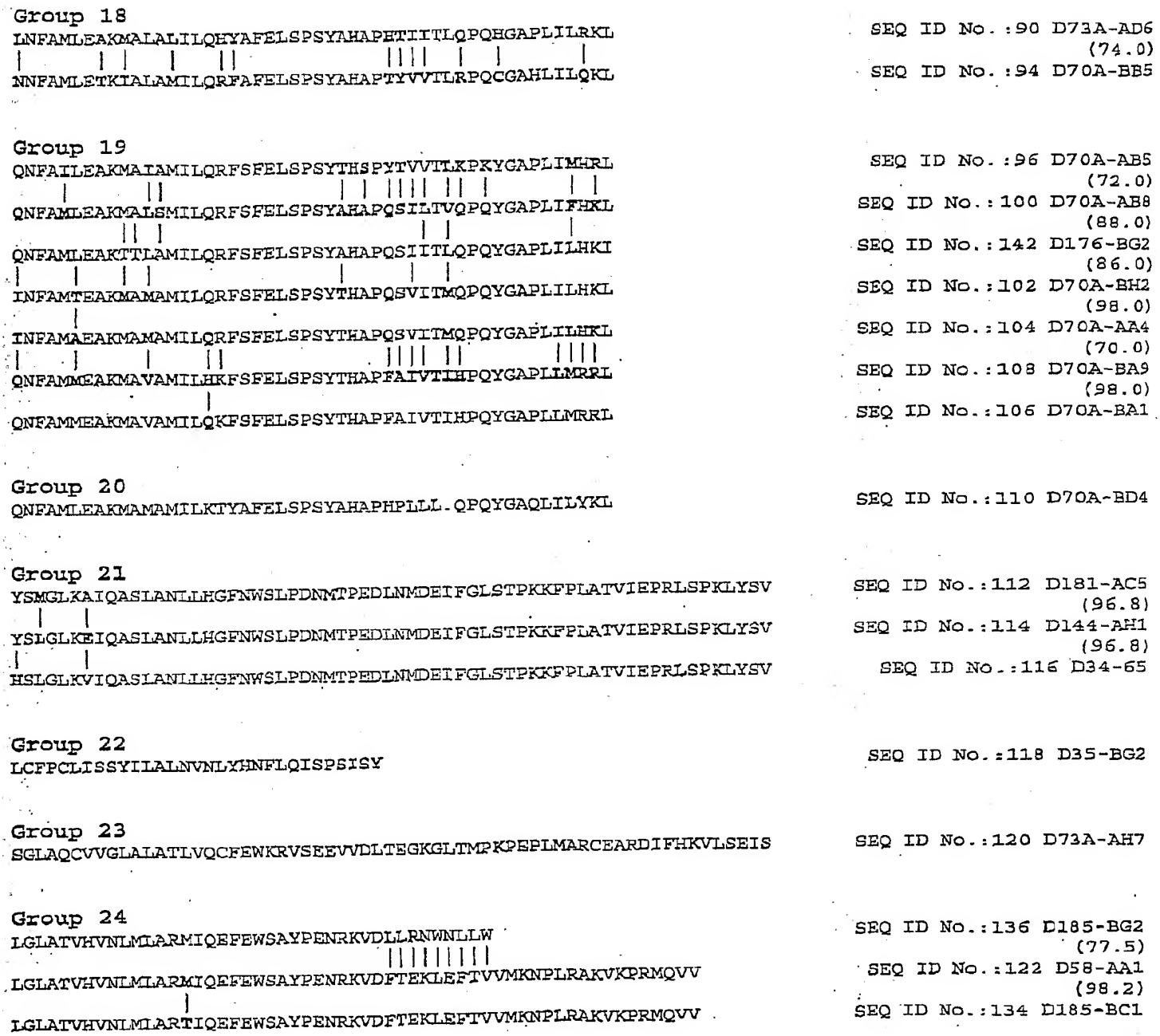


Figure 16 : Amino Acid Identity of Group Members



76

Figure : Amino Acid Identity of Group Members

Group 25
YALAMLHLEYFVANLVWHFRWEAVEGDDVDLSEKLEFTVVMKNPLRARICPRVNSI SEQ ID No.:124 D73A-AE10

Group 26
QQVGLLRTTIFIASLLSEYKLKPRSHQKQVELTDLNPASWLHSIKGELLVDAIPRKKAAF SEQ ID No.:126 D56A-AC12

Group 27
ITFAKFVNELAIARLMFHFDFSLPKGVKHEDLDVEEAAGITVRRKFPLLAVATPCS SEQ ID No.:128 D177-BF7
(98.2)
ITFAKFVNELAIARLMFHFDFSLPKGVKHADLDVEEAAGITVRRKFPLLAVATPCS SEQ ID No.:140 D185-BD2

Group 28
QRYAINHLMLFIALFTALIDFKRHKTDGCDDIAYIPTIAPKDDCKVFLSQRCTRFPSFS SEQ ID No.:130 D73A-AG3

Group 29
MSFGLANLYLPLAQLLYHFDWKLPTGIKPRDLDLTELSGITIARKGDLYLNATPYQPSRE SEQ ID No.:132 D70A-AA12
(80.0)
ISFGLANVYLPLAQLLYHFDWKLPTGINSSDLDMTESSGVTCARKSDLYLTATPYQLSQE SEQ ID No.:186 176-BF2

Group 30 SEQ ID No.:146 D176-BC3
QNFAMLEAKTTLAMILQRFSFELSPSYAHAPQSIITCNPSMVLHLFCIKYSLLLVSSVSFYVKHESKMLRLVELQNGNAFALVHCRLL

Group 31
ADMGLRAVSLALGALIQCFDWQIEEAESLEESYNSRMTMQNKPLKVVCTPREDLGQLLSQL SEQ ID No.:148 D176-BB3

Group 32
MNYSLQVEHLSIAHMIQGFSFATTTNEPLDMKQGVGLTLPKKTDVEVLITPRLPPTLYQY SEQ ID No.:6 D186-AH4

The percentage identity between most related pairs is noted in (0.0%). Each group had at least 70% identity to another group member. Group 19 contained the lowest percentage identity at 70.0%.

FIG. NO 77: COMPARISON OF SEQUENCE GROUPS

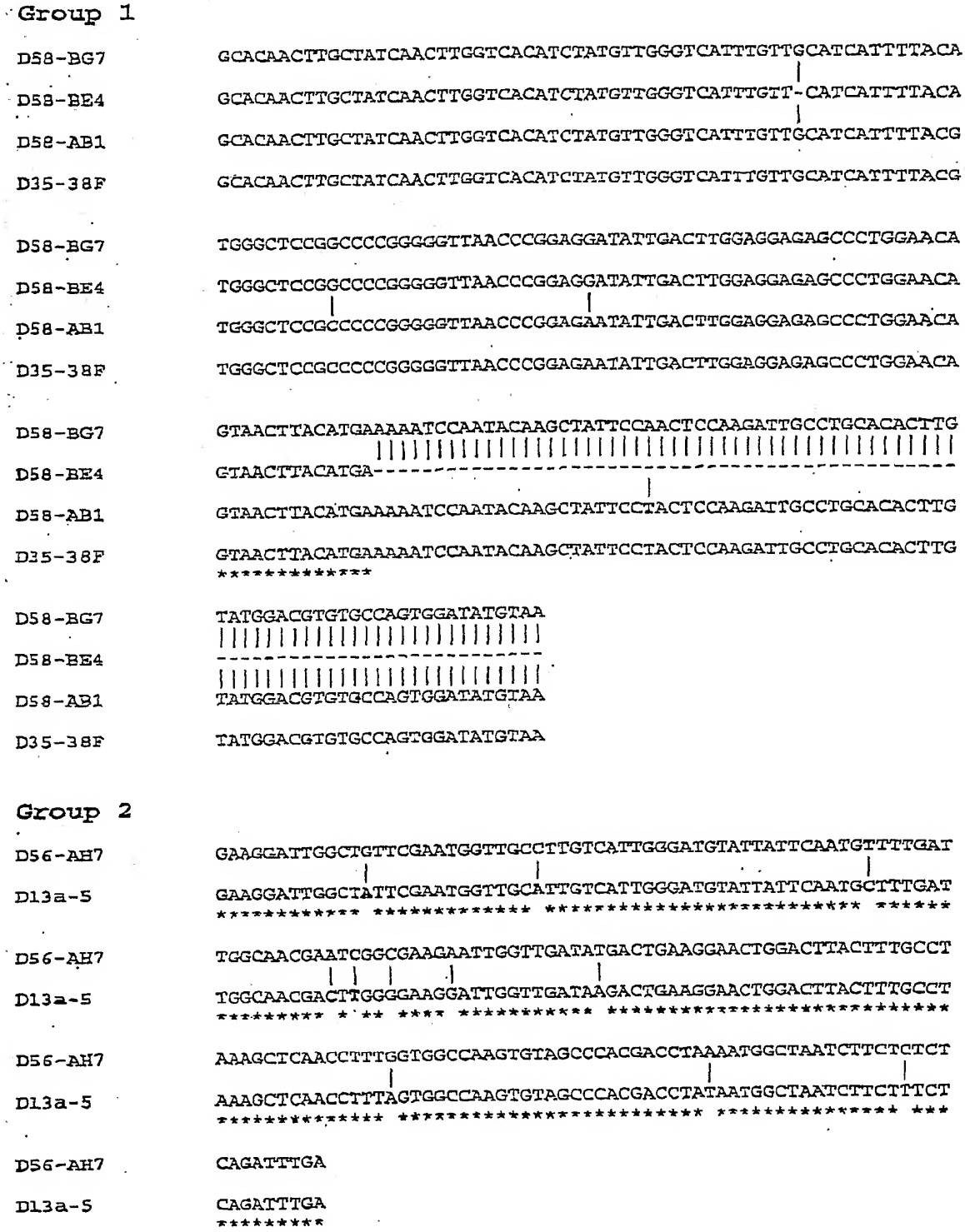
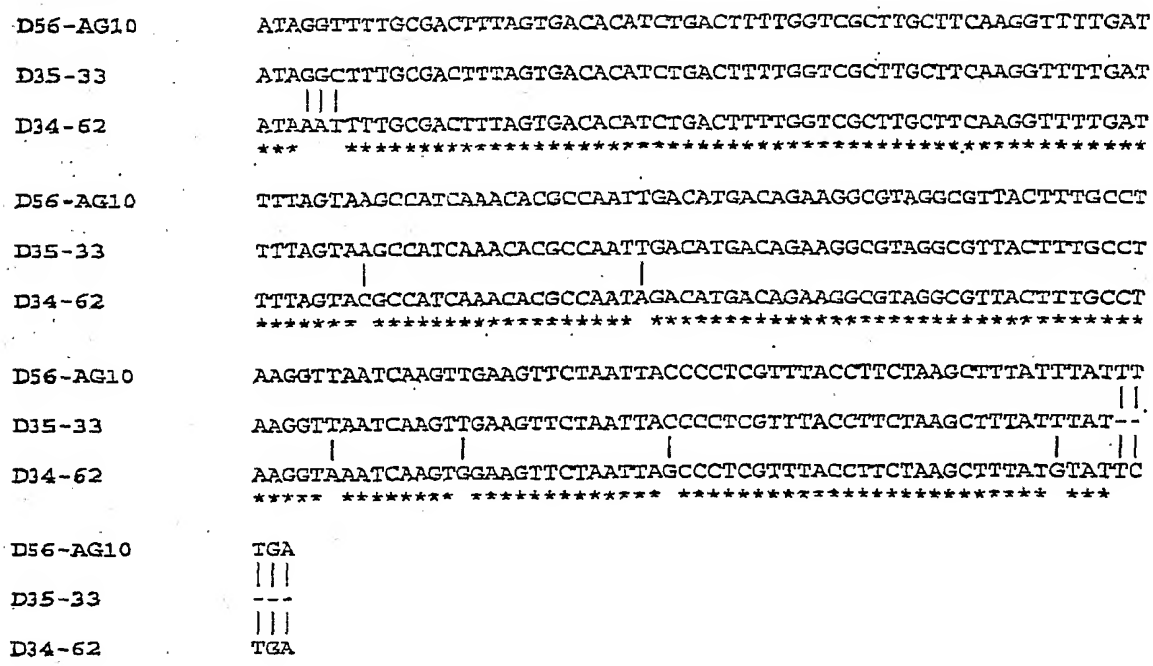


FIG. 7

Group 3



Group 4

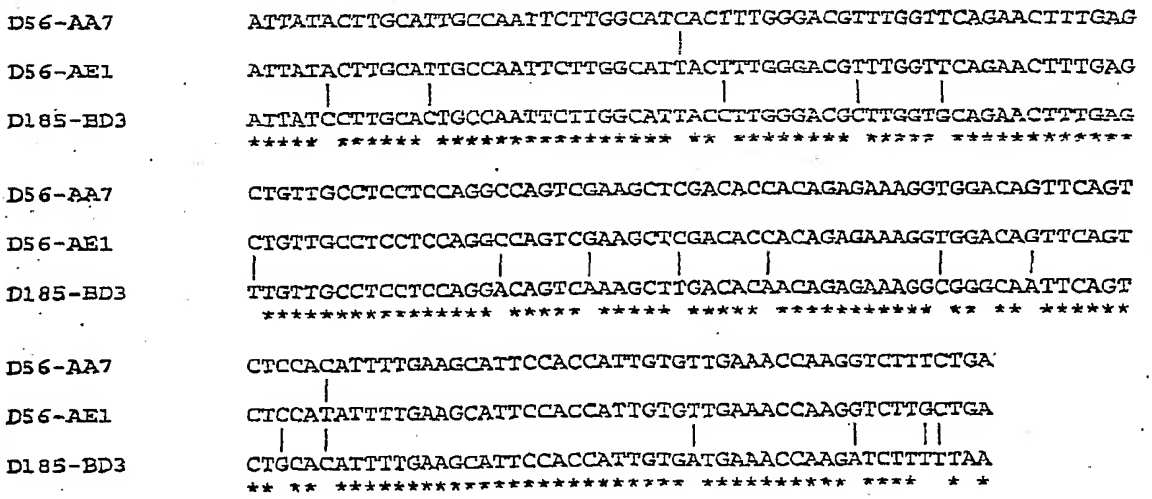


Fig 7J

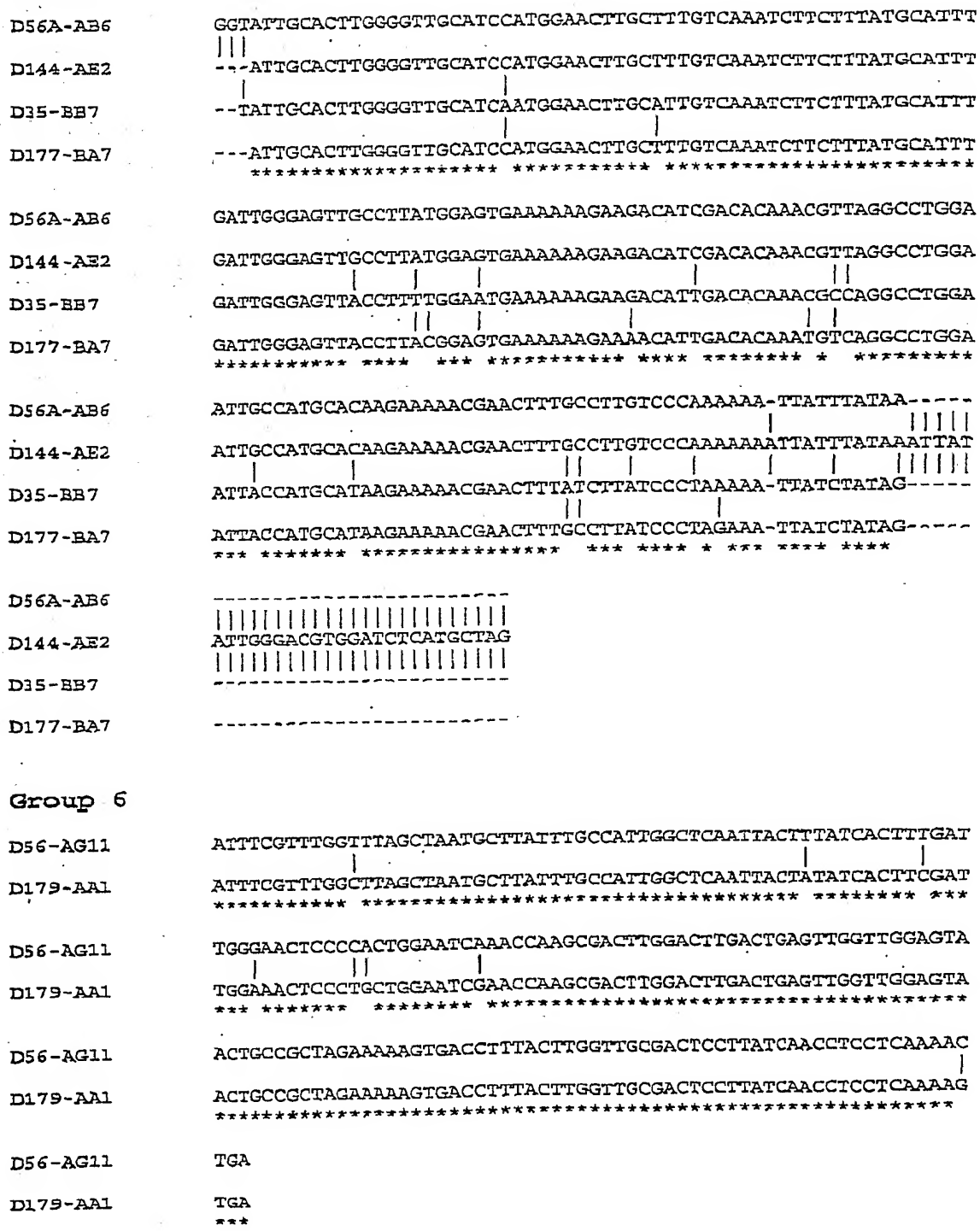

Group 5
D56A-AB6
D144-AE2
D35-BB7
D177-BA7
Group 6
D56-AG11
D179-AA1

Fig. 7J

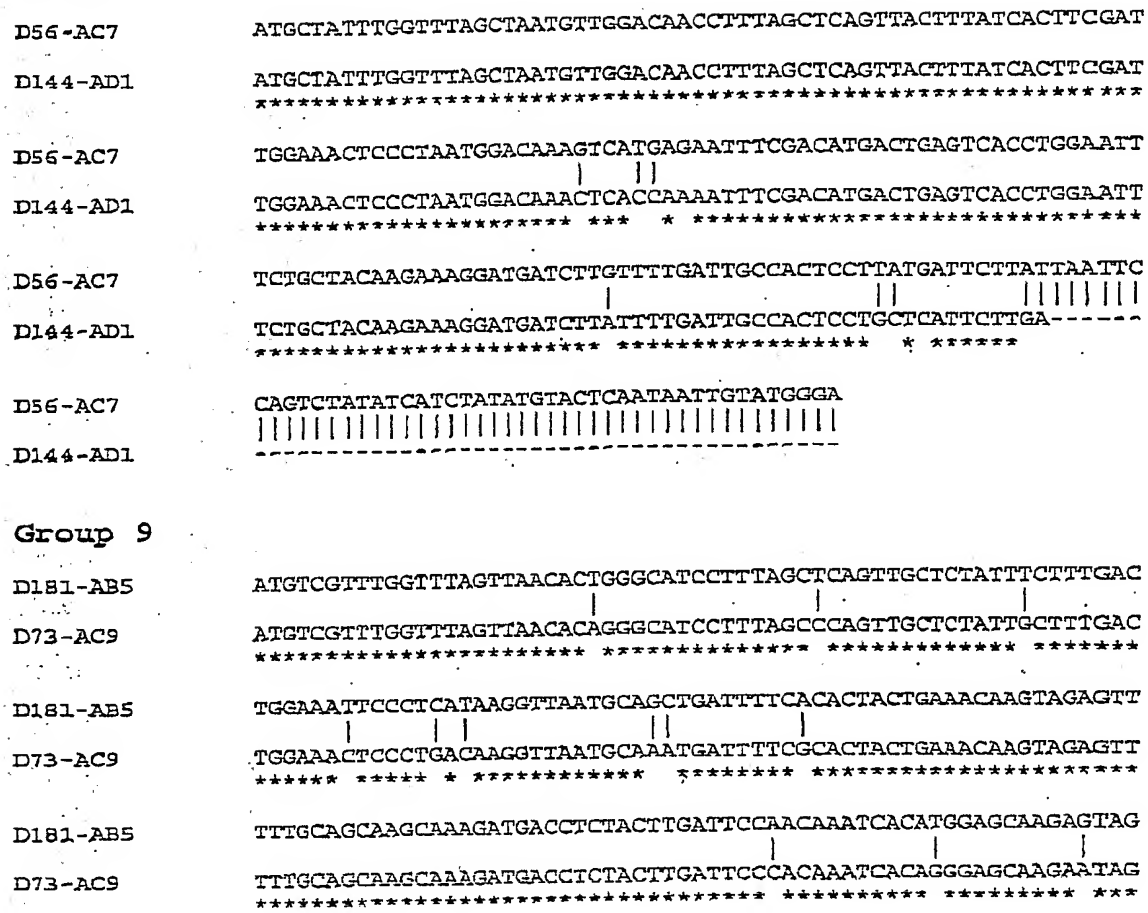

Group 7
D56-AC7
D144-AD1
D56-AC7
D144-AD1
D56-AC7
D144-AD1
D56-AC7
D144-AD1
Group 9
D181-AB5
D73-AC9
D181-AB5
D73-AC9
D181-AB5
D73-AC9

FIG. 7F

Group 11

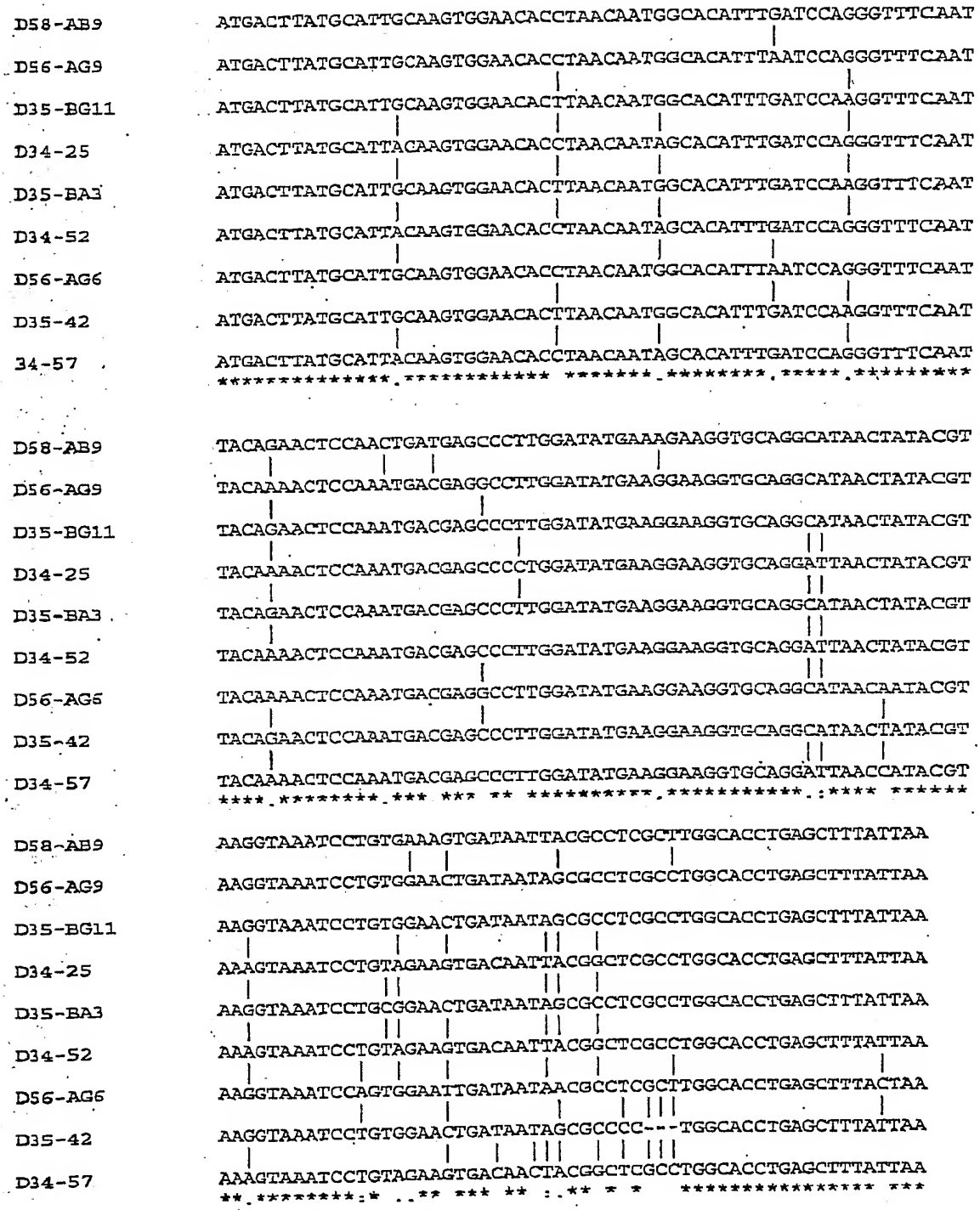


FIG. 77

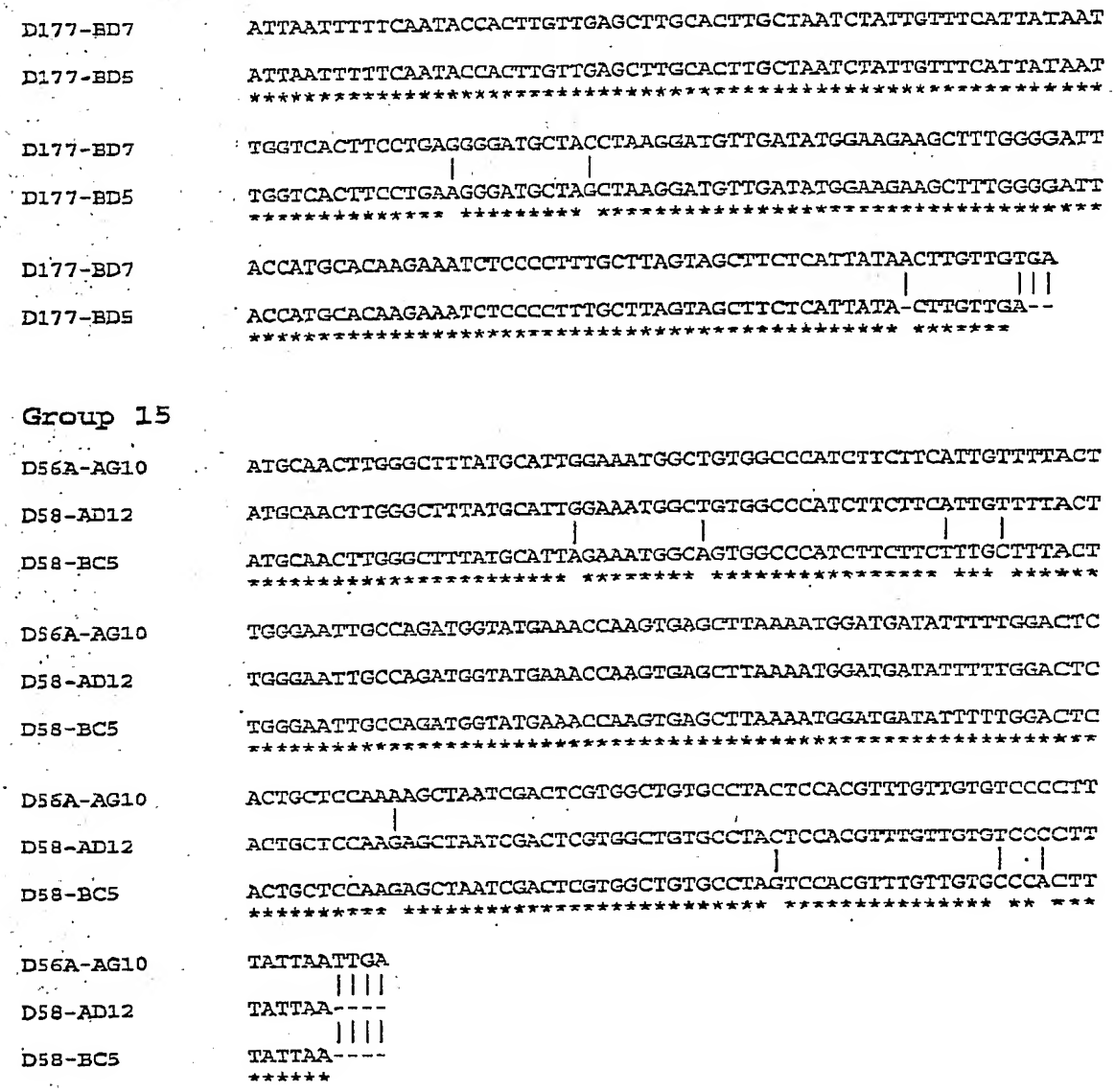

Group 14
D177-BD7 ATTAATTTTTCAATACCACTTGTTGAGCTTGCACTTGCTAATCTATTGTTTCATTATAAT
D177-BD5 ATTAATTTTTCAATACCACTTGTTGAGCTTGCACTTGCTAATCTATTGTTTCATTATAAT
D177-BD7 TGGTCACTTCCTGAGGGGATGCTACCTAAGGATGTTGATATGGAAGAAGCTTTGGGGATT
D177-BD5 TGGTCACTTCCTGAAGGGATGCTAGCTAAGGATGTTGATATGGAAGAAGCTTTGGGGATT
D177-BD7 ACCATGCACAAGAAATCTCCCCTTTGCTTAGTAGCTTCTCATTATAACTTGTTGTGA
D177-BD5 ACCATGCACAAGAAATCTCCCCTTTGCTTAGTAGCTTCTCATTATA-CTTGTTGA--
Group 15
D56A-AG10 ATGCAACTTGGGCTTTATGCATTGGAAATGGCTGTGGCCCATCTTCTTCATTGTTTTACT
D58-AD12 ATGCAACTTGGGCTTTATGCATTGGAAATGGCTGTGGCCCATCTTCTTCATTGTTTTACT
D58-BC5 ATGCAACTTGGGCTTTATGCATTAGAAATGGCAGTGGCCCATCTTCTTCTTTGCTTTACT
D56A-AG10 TGGGAATTGCCAGATGGTATGAAACCAAGTGAGCTTAAAATGGATGATATTTTTGGACTC
D58-AD12 TGGGAATTGCCAGATGGTATGAAACCAAGTGAGCTTAAAATGGATGATATTTTTGGACTC
D58-BC5 TGGGAATTGCCAGATGGTATGAAACCAAGTGAGCTTAAAATGGATGATATTTTTGGACTC
D56A-AG10 ACTGCTCCAAAAGCTAATCGACTCGTGGCTGTGCCTACTCCACGTTTGTTGTGTCCCCTT
D58-AD12 ACTGCTCCAAGAGCTAATCGACTCGTGGCTGTGCCTACTCCACGTTTGTTGTGTCCCCTT
D58-BC5 ACTGCTCCAAGAGCTAATCGACTCGTGGCTGTGCCTAGTCCACGTTTGTTGTGCCCACTT
D56A-AG10 TATTAATTGA
D58-AD12 TATTAA----
D58-BC5 TATTAA----

FIG 7

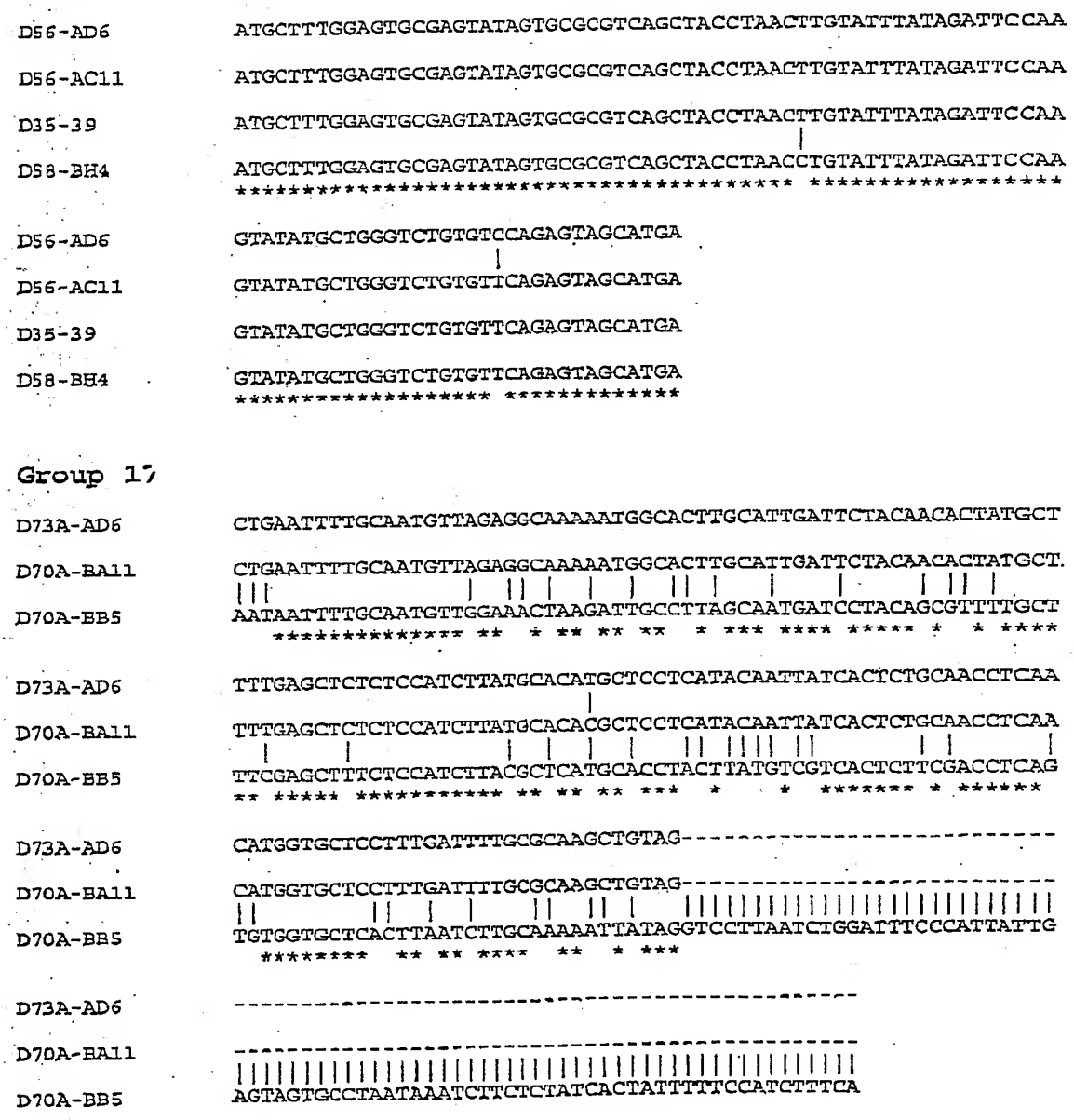

Group 16
D56-AD6
D56-AC11
D35-39
D58-BH4
Group 17
D73A-AD6
D70A-BA11
D70A-BB5

FIG 77

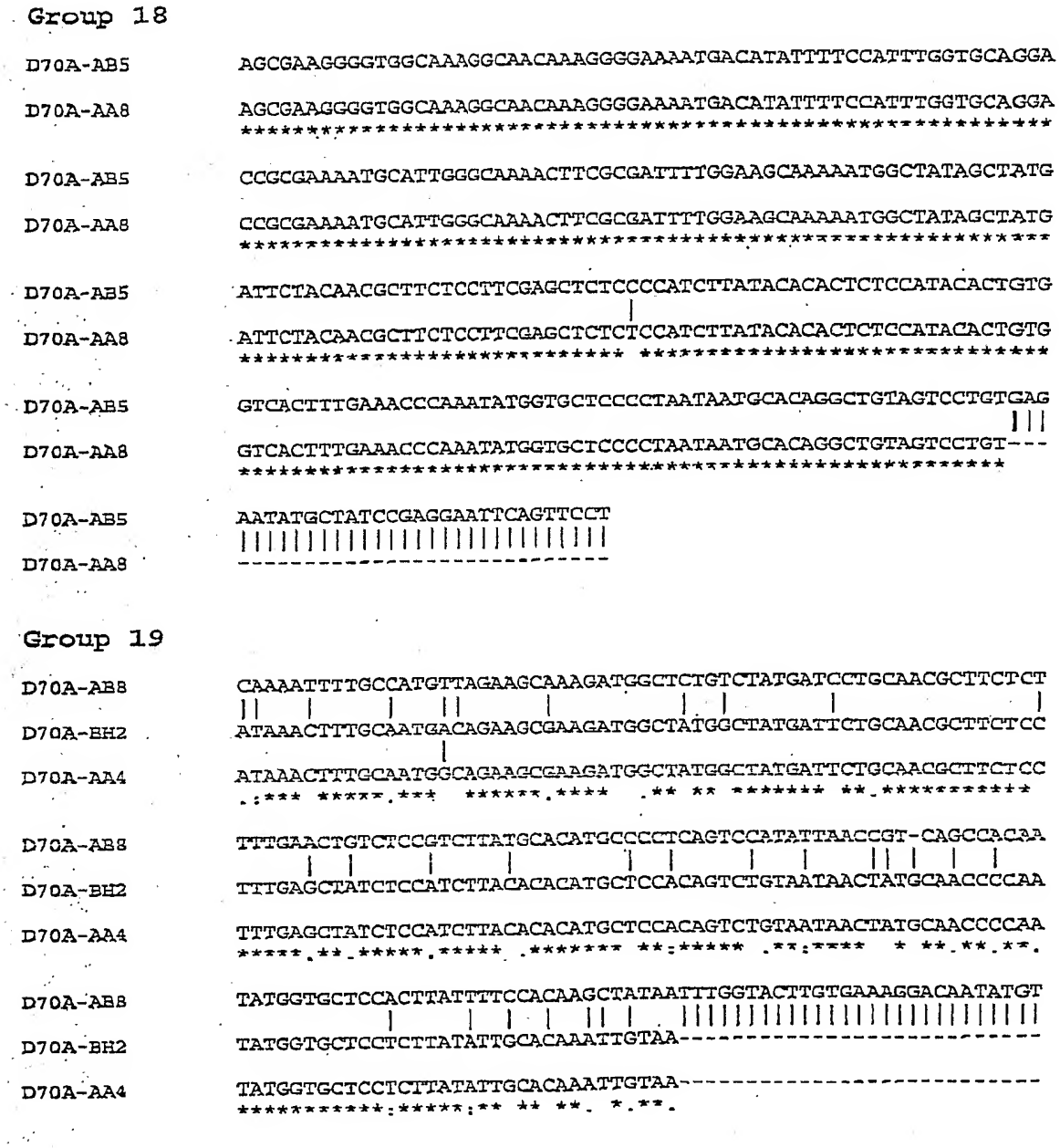

Group 18
D70A-AB5
D70A-AA8
D70A-AB5
D70A-AA8
D70A-AB5
D70A-AA8
D70A-AB5
D70A-AA8
Group 19
D70A-AB8
D70A-BH2
D70A-AA4
D70A-AB8
D70A-BH2
D70A-AA4
D70A-AB8
D70A-BH2
D70A-AA4

FIG 7J

Group 20

```
D70A-BA1    CAAAACTTTGCAATGATGGAAGCAAAAATGGCAGTAGCTATGATACTACAAAAATTTTCC
D70A-BA9    CAAAACTTTGCAATGATGGAAGCAAAAATGGCAGTAGCTATGATACTACATAAATTTTCC
D176-BG2    CAAAATTTTGCCATGTTAGAAGCAAAGACTACTTTGGCTATGATCCTACAACGCTTCTCC
            ***** ***** *** * ******** *  *  * ******** *****    ** ***

D70A-BA1    TTTGAACTATCCCCTTCTTATACACATGCTCCATTTGCAATTGTGACTATTCATCCTCAG
D70A-BA9    TTTGAACTATCCCCTTCTTATACACATGCTCCATTTGCAATTGTGACTATTCATCCTCAG
D176-BG2    TTTCAACTGTCTCCATCTTATGCACATGCTCCTCAGTCCATAATAACTTTGCAACCCCAG
            ******** ** ** ****** **********     * **  * *** * ** ** ***

D70A-BA1    TATGGTGCTCCTCTGCTTATGCGCAGACTTTAA
D70A-BA9    TATGGTGCTCCTCTGCTTATGCGCAGACTTTAA
D176-BG2    TATGGTGCTCCACTTATTTTGCATAAAATATAG
            *********** **  ** ***  * * * **
```

Group 22

```
D144-AH1    TATAGCTTGGGGCTCAAGGAGATTCAAGCTAGCTTAGCTAATCTTCTACATGGATTTAAC
D34-65      CATAGCTTGGGGCTCAAGGTGATTCAAGCTAGCTTAGCTAATCTTCTACATGGATTTAAC
D181-AC5    TATAGCATGGGGCTCAAGGCGATTCAAGCTAGCTTAGCTAATCTTCTACATGGATTTAAC
             *****:************ ****************************************

D144-AH1    TGGTCATTGCCTGATAATATGACTCCTGAGGACCTCAACATGGATGAGATTTTTGGGCTC
D34-65      TGGTCATTGCCTGATAATATGACTCCTGAGGACCTCAACATGGATGAGATTTTTGGGCTC
D181-AC5    TGGTCATTGCCTGATAATATGACTCCTGAGGACCTCAACATGGATGAGATTTTTGGGCTC
            ************************************************************

D144-AH1    TCTACACCTAAAAAATTTCCACTTGCTACTGTGATTGAGCCAAGACTTTCACCAAAACTT
D34-65      TCTACACCTAAAAAATTTCCACTTGCTACTGTGATTGAGCCAAGACTTTCACCAAAACTT
D181-AC5    TCTACACCTAAAAAATTTCCACTTGCTACTGTGATTGAGCCAAGACTTTCACCAAAACTT
            ************************************************************

D144-AH1    TACTCTGTTTGA
D34-65      TACTCTGTTTGA
D181-AC5    TACTCTGTTTGA
            ************
```

FIG. 7J

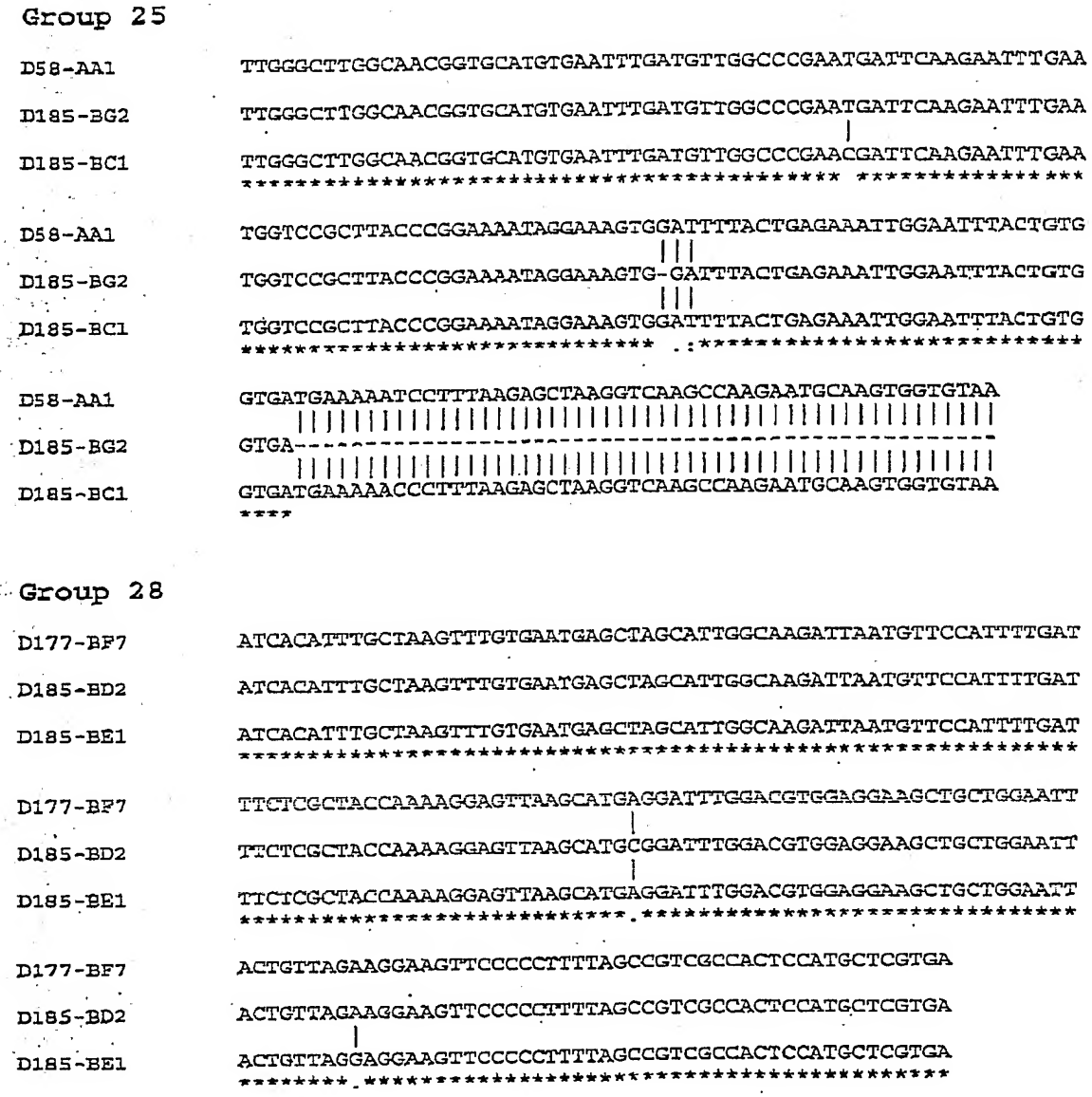
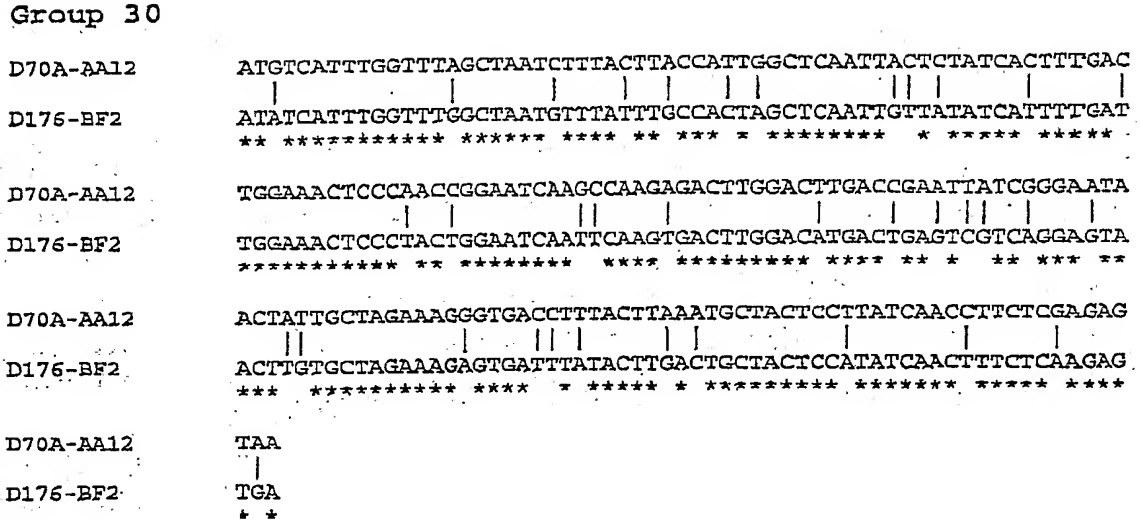

Group 25
D58-AA1
D185-BG2
D185-BC1
Group 28
D177-BF7
D185-BD2
D185-BE1

FIG 77


Group 30
D70A-AA12
D176-BF2
D70A-AA12
D176-BF2
D70A-AA12
D176-BF2
D70A-AA12
D176-BF2
TAA
TGA